Amazon EMR Release Guide

Amazon EMR

Amazon EMR: Amazon EMR Release Guide

Table of Contents

About Amazon EMR Releases

An Amazon EMR release is a set of open-source applications from the big-data ecosystem. Each release comprises different big-data applications, components, and features that you select to have Amazon EMR install and configure when you create a cluster. Applications are packaged using a system based on Apache BigTop, which is an open-source project associated with the Hadoop ecosystem. This guide provides information for applications included in Amazon EMR releases.

For more information about getting started and working with Amazon EMR, see the Amazon EMR Management Guide.

When you launch a cluster, you can choose from multiple releases of Amazon EMR. This allows you to test and use application versions that fit your compatibility requirements. You specify the release number with the *release label*. Release labels are in the form `emr-x.x.x`. For example, `emr-7.2.0`.

Beginning with Amazon EMR 5.18.0, you can use the Amazon EMR artifact repository to build your job code against the exact versions of libraries and dependencies that are available with specific Amazon EMR releases. For more information, see Checking dependencies using the Amazon EMR artifact repository.

To get updates when a new Amazon EMR release is available, subscribe to the RSS feed for Amazon EMR release notes.

Latest release details, including application versions, release notes, components, and configuration classifications of Amazon EMR 7.x, 6.x, and 5.x series:

- Amazon EMR Release 7.2.0
- Amazon EMR Release 6.15.0
- Amazon EMR Release 5.36.2

> **Note**
>
> New Amazon EMR releases are made available in different Regions over a period of several days, beginning with the first Region on the initial release date. The latest release version may not be available in your Region during this period.

Release notes for the latest Amazon EMR releases and a history of all releases:

- What's new?
- Amazon EMR archive of release notes

A comprehensive history of application versions in each Amazon EMR release:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Details for each Amazon EMR release and differences between release series, where applicable:

- Amazon EMR 7.x release versions
- Amazon EMR 6.x release versions
- Amazon EMR 5.x release versions
- Amazon EMR 4.x release versions
- Amazon EMR 2.x and 3.x AMI versions

Amazon EMR standard support

Understanding Amazon EMR releases

The *Amazon EMR release* is the vehicle that delivers the necessary software to run your open-source applications on EC2, EKS, or Serverless platforms. An Amazon EMR release is composed of three pieces: the *Runtime Environment*, *Core Engines*, and *Extras*.

- **Runtime environment –** The runtime environment includes the operating system that's included as part of the Amazon Machine Image (AMI) or other container image when you launch your cluster or container (for example, Amazon Linux 2023). It also includes the language runtimes such as the Amazon Corretto JDK, along with other related tools that the image installs.
- **Core engines –** These include the core open-source software such as Apache Hive, and open table formats such as Apache Iceberg.

- **Extras** – These include convenience libraries and Python packages such as `mariadb-connector-java` and open-source software such as Apache Pig.

The Amazon EMR release version that you select bundles all these components into a new version of Amazon Linux (AL) based Amazon Machine Images (AMI) or container images. Amazon EMR maintains the same AL version for all nodes of a cluster, EKS container, or serverless application. Amazon EMR aims to get the latest *Runtime Environment* within 90 days of their release by Amazon Linux, and new open-source versions of *Core Engines* within 90 days from the upstream release. *Extras* are released as needed. You can find the list of the software in release notes for each Amazon EMR release.

Release versioning

To help you understand the scope of changes in each release, Amazon EMR uses semantic versioning. Semantic versioning follows a format of `<major>.<minor>.<patch>`. As an example, let's consider the 7.0.0 release.

The first digit denotes the major version release, or **Major Release**. Major Releases typically bring substantial changes, improvements, and new features to Runtime Environment or Core Engine that are not backward compatible. For example, the 7.x Major Release uses Amazon Linux 2023 with Amazon Corretto 17 JDK as default. As a result, the release contains several breaking changes compared to 6.x Major Release, which ships on Amazon Linux 2 and with Amazon Corretto 8 JDK as default.

The second digit denotes the minor version release, or **Minor Release**. Minor Releases are non-backward compatible releases that contain incremental changes, improvements, and features to Core Engines and Extras, and new functionality. For example, 6.15 uses Apache Spark 3.4.1, compared to 6.11.1, which ships with Apache Spark 3.3.2.

The third digit denotes a patch version release, or **Patch Release**. Patch Releases are intended to be backward compatible releases that contain fixes and patch updates to *Core Engine* and *Extras*, but no new functionality or OSS version upgrades. For example, 6.11.1 and 6.11.0 both contain the same Apache Spark 3.3.2 OSS version. As of 6.6.x, the latest Patch Release offers the latest *Runtime Environment* patch available. *Runtime Environment* does not affect the Patch Release of Amazon EMR. For example, 6.11.1 launched with AL version 2.0.20240109.0 from December 19, 2023 to January 10, 2024, but with AL version 2.0.20240124.0 from January 11, 2024. Some Amazon EMR on EKS releases allow you to add tags or suffixes to the sematic versions to provide you with alternate functionalities. For example, the `emr-6.15.0-latest` release of Amazon EMR on EKS launches an Amazon Corretto 8 JDK enabled release version, whereas the `emr-6.15.0-java17-latest` release of Amazon EMR on EKS allows you to launch an Amazon Corretto 17 JDK enabled release version.

Support policy

Introduction

Amazon EMR aims to issue new Minor Releases at a minimum of once every 90 days, and to provide support for Minor Releases for a period of 24 months starting from the release date. This support covers *Runtime Environment* and *Core Engines* with their associated dependencies, and doesn't cover *Extras* or their associated dependencies. *Runtime Environment* also includes proprietary platform components and APIs that are needed to start, stop, and operate the clusters and applications. This consistent release schedule ensures a predictable cycle, making it more convenient for you to plan, test, and transition to a version that provides support and security.

What to expect with Standard Support

Standard Support provides fixes on technical support tickets for issues that you encounter for *Runtime Environment* and *Core Engines* components under recommended configurations. All fixes are subject to availability. The recommended configuration is the use of Amazon EMR without any modification, additions, or changes to the binaries and configurations present in an Amazon EMR release, except those found in the Amazon EMR Documentation. Amazon EMR deploys fixes to the latest patch, minor, or major version of the Amazon EMR release within 90 days of us verifying the fix. Amazon EMR automatically applies fixes when you launch a new EMR on EC2 cluster, launch a new Amazon EMR on EKS container, or trigger a new EMR Serverless job. Extra components are provided as a convenience and Amazon EMR does **not** provide fixes related to *Extras*.

Components

Standard Support covers *Runtime Environment* and *Core Engines* components – for example, operating system, language runtimes, and core open-source software like Apache Hive and Apache Iceberg. You can find the full list of supported components for each release in *About Amazon EMR Releases* and Amazon Linux FAQs.

The following list describes the support that we provide for different component types under Standard Support:

- **Runtime Environment components**: *Runtime Environment* components will receive fixes on technical support tickets. Fixes are categorized as (a) critical bugs, (b) critical data corruption issues, and (c) critical security issues. When eligible, Amazon EMR will backport fixes for *Runtime Environment* to older versions. For operational compatibility with the open-source Core Engine components, certain *Runtime Environment* components must remain at specific versions to prevent breaking customer applications. For these components, Amazon EMR is dependent on upstream open-source for availability of fixes. When fixes are available in open source, we will provide the latest stable version within 90 days of them being verified by Amazon EMR.

- **Core Engines components**: *Core Engines* provides the latest versions for many open-source projects, each having hundreds of transitive dependency libraries. While the open-source communities managing these projects attempt to address issues and known Common Vulnerabilities and Exposures (CVE) on a frequent basis, the latest versions might still contain known bugs and CVEs. Amazon EMR is dependent on upstream open-source for availability of fixes and will provide the latest stable version as part of the *Core Engine* components within 90 days of them being verified by Amazon EMR. In certain cases, Amazon EMR might provide a fix for a CVE in one of the *Core Engine* components that needs to be addressed ahead of upstream open source. Amazon EMR also provides you with technical support and fixes on features that are added on top of open source under recommended configurations. We don’t backport fixes for *Core Engine* components to older Patch or Minor versions.
- **Extras components**: Amazon EMR does **not** support *Extras* components. Extra components are open-source projects provided as convenience and Amazon EMR does not provide fixes for issues encountered with them. Any support requests or fixes can be addressed through the open-source community supporting these components.

Standard Support lifecycle

The following describes the milestones in the Standard Support lifecycle:

- **Standard Support**: Amazon EMR releases are eligible for Standard Support 24 months from the date of release. You can create technical support tickets and expect updates for issues that you encounter with these releases.
- **End of Support**: After Standard Support ends, Amazon EMR releases enter End of Support (EoS) stage for 12 months. EoS releases are not eligible for technical support and you won't be able to create any tickets for clusters, containers, or jobs running on these releases. EoS releases won't receive any fixes, patches or updates. EoS releases will be removed from the console, but will continue to be available through the API and AWS CLI. You can always continue to run workloads on EoS releases. We strongly recommend that you migrate to the latest Amazon EMR release

so that you continue to receive security patches, remain eligible for technical support, and can create support tickets when needed.

- **End of Life**: After the End of Support (EoS) term, releases are considered End of Life (EoL). Although you can continue to run EoL clusters, Amazon EMR reserves the right to remove EoL releases from the API and SDK on a case-by-case basis due to security and operational concerns. We strongly recommend that you migrate to the latest Amazon EMR, release because EoL versions can be removed from the API and SDK in exceptional cases.

Bridge Support

Amazon EMR announced this new support policy on July 25, 2024. Under this policy, versions of Amazon EMR released on or before July 24th, 2022 are now designated as End of Support. However, to provide you with additional time to plan and migrate to newer versions, Amazon EMR will offer Bridge Support, equivalent to Standard Support, for these older versions that released within the two-year period before this announcement.

After July 25, 2024, you can refer to the current status and support timelines in the release notes.

The following table shows the support status for all existing Amazon EMR releases at time of announcement of the policy, July 25, 2024:

Releases and supported periods

Amazon EMR release version	Initial release date	Standard support end date	End of support start date	End of life start date
7.2.0	July 25, 2024	July 24, 2026	July 25, 2026	July 25, 2027
7.1.0	April 23, 2024	April 22, 2026	April 23, 2026	April 23, 2027
7.0.0	December 19, 2023	December 18, 2025	December 19, 2025	December 19, 2026
5.36.x and 6.6.x – 6.15.x	May 9, 2022 to November 13, 2023	Bridge support until January 24, 2026	January 25, 2026	January 25, 2026

Amazon EMR release version	Initial release date	Standard support end date	End of support start date	End of life start date
• 6.x series: 6.5.0 and lower • 5.x series: 5.35.0 and lower • 4.x, 3.x, and 2.x series	January 1, 2013 to March 30, 2022	Bridge support until July 24, 2025	July 25, 2025	July 25, 2025

Considerations

Standard Support is available for all Amazon EMR deployment models (EMR on EC2, Amazon EMR on EKS, and EMR Serverless), in all Regions where Amazon EMR is available, at no additional cost. Clusters running with recommended configurations are automatically eligible for support as described in the policy, so you don't need to take any additional actions to activate support.

- Standard Support only supports components required for Amazon EMR clusters. Amazon EMR can't guarantee security patching and fix availability in the case when *Core Engine* open-source components reach EoL upstream, or when security updates are no longer available for dependencies. While you can opt-in to install *Extras*, Amazon EMR won't support them or their dependencies. For example, you can install third-party applications in your custom AMI to harden the security of your cluster, install additional components or copy objects using bootstrap action scripts, or SSH into your cluster and upgrade the default package versions. Amazon EMR does not support these components. Standard Support doesn't cover customer provided bootstrap actions, packages, libraries, your custom code and bring-your-own custom applications that you can configure Amazon EMR to install for your convenience.
- Your existing clusters won't be impacted, regardless of which Amazon EMR release they're running. You can continue to run existing clusters without disruption. You can also continue to launch new clusters and run jobs on any of the existing releases and new releases. All existing releases and new releases at time of the policy becoming effective are covered by Standard Support for 24 months from initial date of release of the Amazon EMR release version. Amazon EMR will provide you with bridge support during the initial announcement of the policy. To

receive uninterrupted support, we advise testing your applications and promptly upgrading to the most recent Amazon EMR release.

- Amazon EMR won't change Standard Support components on existing releases or clusters. However, Amazon EMR reserves the right to honor the upstream End of Life on a case-by-case basis, and remove such components in the new releases, or move the existing release to End of Support (EoS), or End of Life (EoL) status in exceptional cases. We'll notify you of any removal through available channels.
- When a new Region is launched, Amazon EMR will only support Amazon EMR release versions under Standard Support, released in six months prior to the date of of when new Region becomes generally available.
- Amazon EMR won't automatically update your existing clusters to latest versions. However, you can choose to update new clusters to the latest patch versions if desired.
- The cost of your unsupported Amazon EMR usage will count toward your AWS bill. Even if you are using Amazon EMR in an unsupported way, the costs associated with that usage are still part of your overall AWS consumption and will be included in the calculation of your support fees.

For more information, contact AWS Developer Support.

Amazon EMR 7.x release versions

This section contains application versions, release notes, component versions, and configuration classifications available in each Amazon EMR 7.x release version.

When you launch a cluster, you can choose from multiple releases of Amazon EMR. This allows you to test and use application versions that fit your compatibility requirements. You specify the release number with the *release label*. Release labels are in the form `emr-x.x.x`. For example, `emr-7.2.0`.

New Amazon EMR releases are made available in different Regions over a period of several days, beginning with the first Region on the initial release date. The latest release version may not be available in your Region during this period.

For a comprehensive table of application versions in every Amazon EMR 7.x release, see Application versions in Amazon EMR 7.x releases.

Topics

- Amazon EMR release 7.2.0
- Amazon EMR release 7.1.0
- Amazon EMR release 7.0.0

Application versions in Amazon EMR 7.x releases

For a comprehensive table that lists the application versions available in each Amazon EMR 7.x release, open Application versions in Amazon EMR 7.x releases in your browser.

Amazon EMR release 7.2.0

7.2.0 supported lifecycle

The following table describes the supported lifecycle dates for Amazon EMR release 7.2.0.

Support phase	Date	
Initial release date	July 25 2024	
Standard support until	July 24, 2026	
End of support	July 25, 2026	
End of life	July 25, 2027	

For a list of critical and high common vulnerabilities and exposures (CVEs) that don't affect EMR clusters under the recommended software and configurations, see 7.2.0 known common vulnerabilities and exposures for core engines. For a list of CVEs that are fixed in the 7.2.0 release, see 7.2.0 fixed common vulnerabilities and exposures.

7.2.0 common vulnerabilities and exposures

This page lists all of the common vulnerabilities and exposures (CVEs) related to Amazon EMR 7.2.0.

7.2.0 known common vulnerabilities and exposures for core engines

The following table lists all CVEs that don't impact EMR clusters that run on Amazon EMR 7.2.0. Amazon EMR is dependent on upstream open-source for availability of fixes and will provide the latest stable version as part of the *Core Engine* components within 90 days of Amazon EMR verifying the fixes.

This table was updated on September 16th, 2024.

CVE ID	Severity	CVE details URL
CVE-2023-2976	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-2976
CVE-2021-31684	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-31684
CVE-2022-1471	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2022-1471
CVE-2022-25857	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-25857
CVE-2022-3171	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3171
CVE-2022-3509	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3509
CVE-2022-3510	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3510
CVE-2022-40149	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-40149
CVE-2022-40150	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-40150
CVE-2022-40152	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-40152
CVE-2022-42003	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42003

CVE ID	Severity	CVE details URL
CVE-2022-42003	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42003
CVE-2022-42004	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42004
CVE-2022-45685	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-45685
CVE-2022-45693	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-45693
CVE-2023-1370	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-1370
CVE-2023-1436	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-1436
CVE-2023-2976	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-2976
CVE-2023-2976	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-2976
CVE-2023-36478	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-36478
CVE-2023-39410	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-39410
CVE-2023-44981	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2023-44981
CVE-2015-5237	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2015-5237
CVE-2017-15095	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2017-15095

CVE ID	Severity	CVE details URL
CVE-2017-17485	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2017-17485
CVE-2017-7525	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2017-7525
CVE-2018-11307	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-11307
CVE-2018-12022	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2018-12022
CVE-2018-12023	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2018-12023
CVE-2018-14718	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-14718
CVE-2018-14719	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-14719
CVE-2018-14720	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-14720
CVE-2018-14721	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-14721
CVE-2018-19360	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-19360
CVE-2018-19361	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-19361
CVE-2018-19362	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-19362
CVE-2018-5968	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2018-5968

CVE ID	Severity	CVE details URL
CVE-2018-7489	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-7489
CVE-2019-10172	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-10172
CVE-2019-10202	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-10202
CVE-2019-12086	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-12086
CVE-2019-14379	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-14379
CVE-2019-14439	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-14439
CVE-2019-14540	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-14540
CVE-2019-14892	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-14892
CVE-2019-14893	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-14893
CVE-2019-16335	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-16335
CVE-2019-16942	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-16942
CVE-2019-16943	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-16943
CVE-2019-17267	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-17267

CVE ID	Severity	CVE details URL
CVE-2019-17531	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-17531
CVE-2019-20330	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-20330
CVE-2020-10650	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10650
CVE-2020-10672	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10672
CVE-2020-10673	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10673
CVE-2020-10968	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10968
CVE-2020-10969	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10969
CVE-2020-11111	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11111
CVE-2020-11112	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11112
CVE-2020-11113	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11113
CVE-2020-11619	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11619
CVE-2020-11620	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11620
CVE-2020-14060	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-14060

CVE ID	Severity	CVE details URL
CVE-2020-14061	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-14061
CVE-2020-14062	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-14062
CVE-2020-14195	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-14195
CVE-2020-24616	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-24616
CVE-2020-24750	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-24750
CVE-2020-35490	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-35490
CVE-2020-35491	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-35491
CVE-2020-35728	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-35728
CVE-2020-36179	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36179
CVE-2020-36180	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36180
CVE-2020-36181	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36181
CVE-2020-36182	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36182
CVE-2020-36183	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36183

CVE ID	Severity	CVE details URL
CVE-2020-36184	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36184
CVE-2020-36185	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36185
CVE-2020-36186	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36186
CVE-2020-36187	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36187
CVE-2020-36188	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36188
CVE-2020-36189	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36189
CVE-2020-36518	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36518
CVE-2020-8840	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2020-8840
CVE-2020-9546	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2020-9546
CVE-2020-9547	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2020-9547
CVE-2020-9548	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2020-9548
CVE-2021-20190	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-20190
CVE-2021-31684	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-31684

CVE ID	Severity	CVE details URL
CVE-2021-33813	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-33813
CVE-2022-3171	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3171
CVE-2022-3171	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3171
CVE-2022-3509	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3509
CVE-2022-3509	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3509
CVE-2022-3510	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3510
CVE-2022-3510	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3510
CVE-2022-42003	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42003
CVE-2022-42004	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42004
CVE-2023-1370	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-1370
CVE-2023-25194	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-25194
CVE-2023-2976	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-2976
CVE-2023-2976	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-2976

CVE ID	Severity	CVE details URL
CVE-2023-34610	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-34610
CVE-2023-3635	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-3635
CVE-2023-39410	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-39410
CVE-2023-43642	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-43642
CVE-2023-44981	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2023-44981
CVE-2012-0881	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2012-0881
CVE-2013-4002	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2013-4002
CVE-2015-1832	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2015-1832
CVE-2015-5237	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2015-5237
CVE-2018-11307	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-11307
CVE-2018-12022	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2018-12022
CVE-2018-12023	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2018-12023
CVE-2018-14718	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-14718

CVE ID	Severity	CVE details URL
CVE-2018-14719	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-14719
CVE-2018-14720	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-14720
CVE-2018-14721	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-14721
CVE-2018-19360	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-19360
CVE-2018-19361	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-19361
CVE-2018-19362	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-19362
CVE-2018-7489	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-7489
CVE-2019-12086	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-12086
CVE-2019-14379	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-14379
CVE-2019-14439	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-14439
CVE-2019-14540	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-14540
CVE-2019-14892	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-14892
CVE-2019-14893	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-14893

CVE ID	Severity	CVE details URL
CVE-2019-16335	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-16335
CVE-2019-16942	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-16942
CVE-2019-16943	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-16943
CVE-2019-17267	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-17267
CVE-2019-17531	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-17531
CVE-2019-20330	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-20330
CVE-2020-10650	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10650
CVE-2020-10672	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10672
CVE-2020-10673	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10673
CVE-2020-10968	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10968
CVE-2020-10969	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10969
CVE-2020-11111	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11111
CVE-2020-11112	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11112

CVE ID	Severity	CVE details URL
CVE-2020-11113	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11113
CVE-2020-11619	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11619
CVE-2020-11620	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11620
CVE-2020-13936	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-13936
CVE-2020-13949	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-13949
CVE-2020-14060	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-14060
CVE-2020-14061	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-14061
CVE-2020-14062	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-14062
CVE-2020-14195	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-14195
CVE-2020-24616	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-24616
CVE-2020-24750	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-24750
CVE-2020-25649	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-25649
CVE-2020-35490	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-35490

CVE ID	Severity	CVE details URL
CVE-2020-35491	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-35491
CVE-2020-35728	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-35728
CVE-2020-36179	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36179
CVE-2020-36180	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36180
CVE-2020-36181	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36181
CVE-2020-36182	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36182
CVE-2020-36183	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36183
CVE-2020-36184	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36184
CVE-2020-36185	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36185
CVE-2020-36186	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36186
CVE-2020-36187	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36187
CVE-2020-36188	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36188
CVE-2020-36189	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36189

CVE ID	Severity	CVE details URL
CVE-2020-36518	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36518
CVE-2020-8840	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2020-8840
CVE-2020-9546	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2020-9546
CVE-2020-9547	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2020-9547
CVE-2020-9548	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2020-9548
CVE-2021-20190	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-20190
CVE-2021-31684	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-31684
CVE-2021-34538	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-34538
CVE-2021-34538	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-34538
CVE-2021-35515	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-35515
CVE-2021-35516	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-35516
CVE-2021-35517	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-35517
CVE-2021-36090	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-36090

CVE ID	Severity	CVE details URL
CVE-2022-25647	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-25647
CVE-2022-3171	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3171
CVE-2022-3509	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3509
CVE-2022-3510	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3510
CVE-2022-36364	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-36364
CVE-2022-37865	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2022-37865
CVE-2022-37866	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-37866
CVE-2022-39135	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2022-39135
CVE-2022-40149	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-40149
CVE-2022-40150	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-40150
CVE-2022-42003	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42003
CVE-2022-42003	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42003
CVE-2022-42004	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42004

CVE ID	Severity	CVE details URL
CVE-2022-42004	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42004
CVE-2022-45685	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-45685
CVE-2022-45693	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-45693
CVE-2022-46337	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2022-46337
CVE-2022-46751	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-46751
CVE-2023-1370	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-1370
CVE-2023-1436	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-1436
CVE-2023-2976	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-2976
CVE-2023-2976	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-2976
CVE-2023-2976	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-2976
CVE-2023-34453	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-34453
CVE-2023-34454	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-34454
CVE-2023-34455	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-34455

CVE ID	Severity	CVE details URL
CVE-2023-39410	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-39410
CVE-2023-43642	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-43642
CVE-2022-25647	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-25647
CVE-2022-42003	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42003
CVE-2022-42889	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2022-42889
CVE-2022-46751	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-46751
CVE-2023-1370	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-1370
CVE-2023-20883	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-20883
CVE-2023-25194	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-25194
CVE-2023-2976	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-2976
CVE-2023-2976	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-2976
CVE-2023-34453	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-34453
CVE-2023-34454	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-34454

CVE ID	Severity	CVE details URL
CVE-2023-34455	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-34455
CVE-2023-39410	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-39410
CVE-2023-39410	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-39410
CVE-2023-39410	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-39410
CVE-2023-43642	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-43642
CVE-2023-5072	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-5072
CVE-2018-8768	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2018-8768
CVE-2021-32797	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2021-32797
CVE-2021-32798	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2021-32798
CVE-2022-24758	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-24758
CVE-2023-24580	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-24580
CVE-2023-36053	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-36053
CVE-2023-41164	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-41164

CVE ID	Severity	CVE details URL
CVE-2023-43665	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-43665
CVE-2023-46695	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-46695
CVE-2024-24680	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2024-24680
CVE-2016-5018	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2016-5018
CVE-2016-6796	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2016-6796
CVE-2019-0205	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-0205
CVE-2020-13949	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-13949
CVE-2023-2976	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-2976
CVE-2023-43642	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-43642
CVE-2012-0881	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2012-0881
CVE-2013-4002	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2013-4002
CVE-2019-17566	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-17566
CVE-2019-17566	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-17566

CVE ID	Severity	CVE details URL
CVE-2020-11988	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11988
CVE-2021-33813	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-33813
CVE-2022-25647	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-25647
CVE-2022-34169	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-34169
CVE-2022-40146	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-40146
CVE-2022-41704	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-41704
CVE-2022-42890	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42890
CVE-2022-44729	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-44729
CVE-2022-44729	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-44729
CVE-2022-44729	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-44729
CVE-2022-46337	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2022-46337
CVE-2023-4759	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-4759
CVE-2015-2156	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2015-2156

CVE ID	Severity	CVE details URL
CVE-2017-18640	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2017-18640
CVE-2017-18640	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2017-18640
CVE-2019-16869	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-16869
CVE-2019-20444	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-20444
CVE-2019-20445	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-20445
CVE-2021-37136	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-37136
CVE-2021-37137	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-37137
CVE-2022-1471	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2022-1471
CVE-2022-1471	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2022-1471
CVE-2022-25857	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-25857
CVE-2022-25857	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-25857
CVE-2022-3171	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3171
CVE-2022-3509	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3509

CVE ID	Severity	CVE details URL
CVE-2022-3510	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3510
CVE-2022-42003	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42003
CVE-2022-42004	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42004
CVE-2023-2976	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-2976
CVE-2023-2976	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-2976
CVE-2023-36478	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-36478
CVE-2023-44981	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2023-44981
CVE-2023-44981	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2023-44981
CVE-2015-2080	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2015-2080
CVE-2015-4852	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2015-4852
CVE-2015-6420	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2015-6420
CVE-2015-7501	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2015-7501
CVE-2017-1000034	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2017-1000034

CVE ID	Severity	CVE details URL
CVE-2017-12612	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2017-12612
CVE-2017-15288	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2017-15288
CVE-2017-5637	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2017-5637
CVE-2017-7656	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2017-7656
CVE-2017-7657	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2017-7657
CVE-2017-7658	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2017-7658
CVE-2017-7687	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2017-7687
CVE-2017-9735	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2017-9735
CVE-2017-9790	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2017-9790
CVE-2018-11793	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2018-11793
CVE-2018-11804	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2018-11804
CVE-2018-17190	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-17190
CVE-2018-8012	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2018-8012

CVE ID	Severity	CVE details URL
CVE-2019-0204	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-0204
CVE-2019-10099	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-10099
CVE-2020-17533	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-17533
CVE-2020-9480	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2020-9480
CVE-2021-28165	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-28165
CVE-2021-35515	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-35515
CVE-2021-35515	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-35515
CVE-2021-35516	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-35516
CVE-2021-35516	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-35516
CVE-2021-35517	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-35517
CVE-2021-35517	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-35517
CVE-2021-36090	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-36090
CVE-2021-36090	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-36090

CVE ID	Severity	CVE details URL
CVE-2022-25647	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-25647
CVE-2022-3171	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3171
CVE-2022-3509	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3509
CVE-2022-3510	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3510
CVE-2022-37866	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-37866
CVE-2022-46751	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-46751
CVE-2023-2976	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-2976
CVE-2023-34453	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-34453
CVE-2023-34454	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-34454
CVE-2023-34455	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-34455
CVE-2023-43642	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-43642
CVE-2023-44981	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2023-44981
CVE-2014-0114	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2014-0114

CVE ID	Severity	CVE details URL
CVE-2016-1000338	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2016-1000338
CVE-2016-1000340	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2016-1000340
CVE-2016-1000342	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2016-1000342
CVE-2016-1000343	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2016-1000343
CVE-2016-1000344	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2016-1000344
CVE-2016-1000352	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2016-1000352
CVE-2017-1000487	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2017-1000487
CVE-2017-12629	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2017-12629
CVE-2018-10936	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2018-10936
CVE-2018-3258	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2018-3258
CVE-2018-3827	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2018-3827
CVE-2018-3831	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2018-3831
CVE-2019-0205	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-0205

CVE ID	Severity	CVE details URL
CVE-2019-10086	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-10086
CVE-2019-10086	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-10086
CVE-2019-7611	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-7611
CVE-2020-13692	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-13692
CVE-2020-13949	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-13949
CVE-2020-28491	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-28491
CVE-2020-7692	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2020-7692
CVE-2021-22573	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-22573
CVE-2021-22573	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-22573
CVE-2021-25646	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-25646
CVE-2021-26291	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2021-26291
CVE-2021-26291	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2021-26291
CVE-2021-26919	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-26919

CVE ID	Severity	CVE details URL
CVE-2021-35515	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-35515
CVE-2021-35516	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-35516
CVE-2021-35517	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-35517
CVE-2021-36090	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-36090
CVE-2022-1471	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2022-1471
CVE-2022-25647	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-25647
CVE-2022-25857	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-25857
CVE-2022-31197	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-31197
CVE-2022-3171	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3171
CVE-2022-3171	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3171
CVE-2022-33980	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2022-33980
CVE-2022-3509	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3509
CVE-2022-3509	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3509

CVE ID	Severity	CVE details URL
CVE-2022-3510	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3510
CVE-2022-3510	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3510
CVE-2022-42003	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42003
CVE-2022-42004	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42004
CVE-2022-4244	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-4244
CVE-2022-45688	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-45688
CVE-2023-1370	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-1370
CVE-2023-1428	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-1428
CVE-2023-25194	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-25194
CVE-2023-2976	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-2976
CVE-2023-2976	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-2976
CVE-2023-2976	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-2976
CVE-2023-31418	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-31418

CVE ID	Severity	CVE details URL
CVE-2023-32731	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-32731
CVE-2023-34453	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-34453
CVE-2023-34454	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-34454
CVE-2023-34455	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-34455
CVE-2023-3635	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-3635
CVE-2023-38647	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2023-38647
CVE-2023-39410	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-39410
CVE-2023-43642	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-43642
CVE-2023-44981	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2023-44981
CVE-2023-46120	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-46120
CVE-2023-5072	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-5072
CVE-2023-6378	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-6378
CVE-2023-6481	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-6481

CVE ID	Severity	CVE details URL
CVE-2024-1597	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2024-1597
CVE-2024-1597	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2024-1597
CVE-2021-35958	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2021-35958
CVE-2019-0205	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-0205
CVE-2020-13949	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-13949
CVE-2021-34538	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-34538
CVE-2022-45688	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-45688
CVE-2022-46337	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2022-46337
CVE-2023-2976	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-2976
CVE-2023-3635	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-3635
CVE-2023-5072	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-5072
CVE-2024-21634	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2024-21634
CVE-2021-35515	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-35515

CVE ID	Severity	CVE details URL
CVE-2021-35516	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-35516
CVE-2021-35517	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-35517
CVE-2021-36090	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-36090
CVE-2022-40149	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-40149
CVE-2022-40150	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-40150
CVE-2022-45685	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-45685
CVE-2022-45693	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-45693
CVE-2023-1436	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-1436
CVE-2014-0114	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2014-0114
CVE-2015-2156	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2015-2156
CVE-2018-10054	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2018-10054
CVE-2019-0205	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-0205
CVE-2019-10086	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-10086

CVE ID	Severity	CVE details URL
CVE-2019-10086	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-10086
CVE-2019-16869	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-16869
CVE-2019-17195	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-17195
CVE-2019-20444	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-20444
CVE-2019-20445	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-20445
CVE-2020-13949	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-13949
CVE-2020-9492	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-9492
CVE-2021-37136	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-37136
CVE-2021-37137	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-37137
CVE-2021-37404	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2021-37404
CVE-2022-25168	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2022-25168
CVE-2022-25647	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-25647
CVE-2022-25647	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-25647

CVE ID	Severity	CVE details URL
CVE-2022-26612	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2022-26612
CVE-2022-3171	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3171
CVE-2022-3509	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3509
CVE-2022-3510	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-3510
CVE-2022-40152	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-40152
CVE-2023-2976	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-2976
CVE-2023-2976	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-2976
CVE-2023-34453	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-34453
CVE-2023-34454	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-34454
CVE-2023-34455	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-34455
CVE-2023-3635	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-3635
CVE-2023-3635	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-3635
CVE-2023-38647	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2023-38647

CVE ID	Severity	CVE details URL
CVE-2023-43642	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-43642
CVE-2023-6378	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-6378
CVE-2023-6481	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-6481
CVE-2024-1597	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2024-1597
CVE-2017-15095	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2017-15095
CVE-2017-17485	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2017-17485
CVE-2018-11307	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-11307
CVE-2018-12022	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2018-12022
CVE-2018-12023	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2018-12023
CVE-2018-14718	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-14718
CVE-2018-14719	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-14719
CVE-2018-14720	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-14720
CVE-2018-14721	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-14721

CVE ID	Severity	CVE details URL
CVE-2018-19360	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-19360
CVE-2018-19361	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-19361
CVE-2018-19362	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-19362
CVE-2018-5968	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2018-5968
CVE-2018-7489	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2018-7489
CVE-2019-12086	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-12086
CVE-2019-12086	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-12086
CVE-2019-14379	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-14379
CVE-2019-14379	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-14379
CVE-2019-14439	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-14439
CVE-2019-14439	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-14439
CVE-2019-14540	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-14540
CVE-2019-14540	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-14540

CVE ID	Severity	CVE details URL
CVE-2019-14892	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-14892
CVE-2019-14892	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-14892
CVE-2019-14893	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-14893
CVE-2019-16335	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-16335
CVE-2019-16335	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-16335
CVE-2019-16942	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-16942
CVE-2019-16942	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-16942
CVE-2019-16943	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-16943
CVE-2019-16943	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-16943
CVE-2019-17195	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-17195
CVE-2019-17267	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-17267
CVE-2019-17267	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-17267
CVE-2019-17531	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-17531

CVE ID	Severity	CVE details URL
CVE-2019-17531	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-17531
CVE-2019-20330	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-20330
CVE-2019-20330	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2019-20330
CVE-2020-10650	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10650
CVE-2020-10650	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10650
CVE-2020-10672	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10672
CVE-2020-10672	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10672
CVE-2020-10673	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10673
CVE-2020-10673	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10673
CVE-2020-10968	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10968
CVE-2020-10968	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10968
CVE-2020-10969	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10969
CVE-2020-10969	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10969

CVE ID	Severity	CVE details URL
CVE-2020-11111	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11111
CVE-2020-11111	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11111
CVE-2020-11112	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11112
CVE-2020-11113	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11113
CVE-2020-11619	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11619
CVE-2020-11620	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11620
CVE-2020-11620	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11620
CVE-2020-14060	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-14060
CVE-2020-14060	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-14060
CVE-2020-14061	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-14061
CVE-2020-14061	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-14061
CVE-2020-14062	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-14062
CVE-2020-14062	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-14062

CVE ID	Severity	CVE details URL
CVE-2020-14195	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-14195
CVE-2020-14195	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-14195
CVE-2020-24616	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-24616
CVE-2020-24616	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-24616
CVE-2020-24750	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-24750
CVE-2020-24750	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-24750
CVE-2020-25649	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-25649
CVE-2020-25649	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-25649
CVE-2020-25649	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-25649
CVE-2020-35211	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-35211
CVE-2020-35490	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-35490
CVE-2020-35490	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-35490
CVE-2020-35490	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-35490

CVE ID	Severity	CVE details URL
CVE-2020-35491	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-35491
CVE-2020-35491	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-35491
CVE-2020-35491	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-35491
CVE-2020-35728	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-35728
CVE-2020-35728	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-35728
CVE-2020-35728	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-35728
CVE-2020-36179	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36179
CVE-2020-36179	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36179
CVE-2020-36179	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36179
CVE-2020-36180	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36180
CVE-2020-36180	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36180
CVE-2020-36180	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36180
CVE-2020-36181	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36181

CVE ID	Severity	CVE details URL
CVE-2020-36181	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36181
CVE-2020-36181	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36181
CVE-2020-36182	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36182
CVE-2020-36182	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36182
CVE-2020-36182	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36182
CVE-2020-36183	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36183
CVE-2020-36183	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36183
CVE-2020-36183	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36183
CVE-2020-36184	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36184
CVE-2020-36184	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36184
CVE-2020-36184	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36184
CVE-2020-36185	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36185
CVE-2020-36185	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36185

CVE ID	Severity	CVE details URL
CVE-2020-36185	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36185
CVE-2020-36186	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36186
CVE-2020-36186	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36186
CVE-2020-36186	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36186
CVE-2020-36187	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36187
CVE-2020-36187	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36187
CVE-2020-36187	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36187
CVE-2020-36188	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36188
CVE-2020-36188	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36188
CVE-2020-36188	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36188
CVE-2020-36189	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36189
CVE-2020-36189	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36189
CVE-2020-36189	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36189

CVE ID	Severity	CVE details URL
CVE-2020-36518	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36518
CVE-2020-36518	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36518
CVE-2020-36518	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36518
CVE-2020-36518	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36518
CVE-2020-7692	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2020-7692
CVE-2020-8840	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2020-8840
CVE-2020-8840	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2020-8840
CVE-2020-9546	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2020-9546
CVE-2020-9546	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2020-9546
CVE-2020-9547	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2020-9547
CVE-2020-9547	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2020-9547
CVE-2020-9548	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2020-9548
CVE-2020-9548	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2020-9548

CVE ID	Severity	CVE details URL
CVE-2021-20190	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-20190
CVE-2021-20190	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-20190
CVE-2021-20190	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-20190
CVE-2021-22573	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-22573
CVE-2021-33036	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-33036
CVE-2021-35515	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-35515
CVE-2021-35515	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-35515
CVE-2021-35516	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-35516
CVE-2021-35516	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-35516
CVE-2021-37714	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-37714
CVE-2021-37714	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-37714
CVE-2021-37714	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-37714
CVE-2021-37714	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-37714

CVE ID	Severity	CVE details URL
CVE-2021-46877	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-46877
CVE-2022-25647	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-25647
CVE-2022-32532	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2022-32532
CVE-2022-40149	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-40149
CVE-2022-40150	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-40150
CVE-2022-40664	CRITICAL	https://nvd.nist.gov/vuln/detail/CVE-2022-40664
CVE-2022-42003	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42003
CVE-2022-42003	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42003
CVE-2022-42003	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42003
CVE-2022-42003	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42003
CVE-2022-42004	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42004
CVE-2022-42004	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42004
CVE-2022-42004	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42004

CVE ID	Severity	CVE details URL
CVE-2022-42004	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42004
CVE-2022-45685	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-45685
CVE-2022-45688	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-45688
CVE-2022-45693	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-45693
CVE-2023-1436	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-1436
CVE-2023-2976	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-2976
CVE-2023-2976	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-2976
CVE-2022-40149	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-40149
CVE-2022-40150	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-40150
CVE-2022-45685	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-45685
CVE-2022-45693	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-45693
CVE-2023-1436	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-1436
CVE-2023-6378	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-6378

CVE ID	Severity	CVE details URL
CVE-2023-6378	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-6378
CVE-2023-6481	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-6481
CVE-2023-6481	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-6481

7.2.0 fixed common vulnerabilities and exposures for core engines

The following table lists all fixed CVEs in Amazon EMR 7.2.0.

CVE ID	Severity	CVE details URL
CVE-2019-12402	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-12402
CVE-2021-41561	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-41561

7.2.0 release components

Amazon EMR 7.2.0 supports the following applications:

Core engines:

AmazonCloudWatchAgent, Delta. Flink, Hadoop, HBase, HCatalog, Hive, Hudi, Livy, Iceberg, JupyterEnterpriseGateway, Phoenix, Spark, Tez, Trino, Zookeeper

Extras:

Hue, JupyterHub, MXNet, Oozie, Pig, Presto, Sqoop, TensorFlow, Zeppelin

For more details about the supported packages, see 7.2.0 release component details. For more information about core engines and extras, see Amazon EMR Standard Support.

7.2.0 release components details

See the following table for more information about the *Extras* packages in Amazon EMR 7.2.0.

Application	Application Type	Package	Package Type	Package Version
AmazonCloudWatchAgent	Core	hadoop-hdfs-journalnode	Extras	3.3.6.amzn.4-1.amzn2023
AmazonCloudWatchAgent	Core	ranger-kms	Extras	2.0.0-1.amzn2023
AmazonCloudWatchAgent	Core	s3-dist-cp	Extras	2.32.0-1.amzn2023
AmazonCloudWatchAgent	Core	zookeeper	Extras	3.9.1.amzn.1-1.amzn2023
AmazonCloudWatchAgent	Core	zookeeper-server	Extras	3.9.1.amzn.1-1.amzn2023
Flink	Core	hadoop-client	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Core	hadoop-hdfs	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Core	hadoop-hdfs-datanode	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Core	hadoop-hdfs-journalnode	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Core	hadoop-hdfs-namenode	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Core	hadoop-httpfs	Extras	3.3.6.amzn.4-1.amzn2023

Application	Application Type	Package	Package Type	Package Version
Flink	Core	hadoop-kms	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Core	hadoop-mapreduce	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Core	hadoop-mapreduce-historyserver	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Core	hadoop-yarn	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Core	hadoop-yarn-nodemanager	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Core	hadoop-yarn-proxyserver	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Core	hadoop-yarn-resourcemanager	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Core	hadoop-yarn-timelineserver	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Core	hudi	Extras	0.14.1.amzn.1-1.amzn2023
Flink	Core	ranger-kms	Extras	2.0.0-1.amzn2023
Flink	Core	hadoop	Extras	3.3.6.amzn.4-1.amzn2023

Application	Application Type	Package	Package Type	Package Version
HBase	Core	s3-dist-cp	Extras	2.32.0-1.amzn2023
HCatalog	Core	hadoop-yarn-proxyserver	Extras	3.3.6.amzn.4-1.amzn2023
HCatalog	Core	mariadb-connector-java	Extras	2.7.2-1
Hive	Core	hadoop-httpfs	Extras	3.3.6.amzn.4-1.amzn2023
Hive	Core	hadoop-yarn-proxyserver	Extras	3.3.6.amzn.4-1.amzn2023
Hive	Core	hive-hbase	Extras	3.1.3.amzn.11-1.amzn2023
Hive	Core	mariadb-connector-java	Extras	2.7.2-1
Hive	Core	s3-dist-cp	Extras	2.32.0-1.amzn2023
Phoenix	Core	s3-dist-cp	Extras	2.32.0-1.amzn2023
Spark	Core	spark-datanucleus	Extras	3.5.1.amzn.0-1.amzn2023
Tez	Core	hadoop-mapreduce	Extras	3.3.6.amzn.4-1.amzn2023

Application	Application Type	Package	Package Type	Package Version
Tez	Core	hadoop-mapreduce-historyserver	Extras	3.3.6.amzn.4-1.amzn2023
Tez	Core	hadoop-yarn-proxyserver	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Core	bigtop-utils	Extras	1.2.0-1.amzn2023
Trino	Core	hadoop-client	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Core	hadoop-hdfs	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Core	hadoop-hdfs-datanode	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Core	hadoop-hdfs-journalnode	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Core	hadoop-hdfs-namenode	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Core	hadoop-kms	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Core	hadoop-mapreduce	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Core	hadoop-mapreduce-historyserver	Extras	3.3.6.amzn.4-1.amzn2023

Application	Application Type	Package	Package Type	Package Version
Trino	Core	hadoop-yarn	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Core	hadoop-yarn-nodemanager	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Core	hadoop-yarn-proxyserver	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Core	hadoop-yarn-resourcemanager	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Core	hadoop-yarn-timelineserver	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Core	hive	Extras	3.1.3.amzn.11-1.amzn2023
Trino	Core	hive-hcatalog-server	Extras	3.1.3.amzn.11-1.amzn2023
Trino	Core	mariadb-connector-java	Extras	2.7.2-1
Trino	Core	ranger-kms	Extras	2.0.0-1.amzn2023
Trino	Core	zookeeper	Extras	3.9.1.amzn.1-1.amzn2023
Trino	Core	zookeeper-server	Extras	3.9.1.amzn.1-1.amzn2023

Application	Application Type	Package	Package Type	Package Version
Trino	Core	bigtop-groovy	Extras	2.5.4-1.amzn2023
Trino	Core	bigtop-jsvc	Extras	1.2.4-1.amzn2023
Trino	Core	hadoop	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Core	hadoop-lzo	Extras	0.4.19-1.amzn2023
Trino	Core	hive-hcatalog	Extras	3.1.3.amzn.11-1.amzn2023
Trino	Core	hive-jdbc	Extras	3.1.3.amzn.11-1.amzn2023
Zookeeper	Core	emrfs	Extras	2.63.0-1.amzn2023
Zookeeper	Core	hadoop-client	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Core	hadoop-hdfs	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Core	hadoop-hdfs-datanode	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Core	hadoop-hdfs-journalnode	Extras	3.3.6.amzn.4-1.amzn2023

Application	Application Type	Package	Package Type	Package Version
Zookeeper	Core	hadoop-hdfs-namenode	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Core	hadoop-httpfs	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Core	hadoop-kms	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Core	hadoop-mapreduce	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Core	hadoop-mapreduce-historyserver	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Core	hadoop-yarn	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Core	hadoop-yarn-nodemanager	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Core	hadoop-yarn-proxyserver	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Core	hadoop-yarn-resourcemanager	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Core	hadoop-yarn-timelineserver	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Core	ranger-kms	Extras	2.0.0-1.amzn2023

7.2.0 application versions

The following applications are supported in this release: AmazonCloudWatchAgent, Delta, Flink, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, Iceberg, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Trino, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-7.2.0	emr-7.1.0	emr-7.0.0
AWS SDK for Java	2.23.18, 1.12.705	2.23.18, 1.12.656	2.20.160-amzn-0, 1.12.569
Python	3.9	3.9	3.9
Scala	2.12.17	2.12.17	2.12.17
AmazonCloudWatchAgent	1.300032.2	1.300032.2	1.300031.1
Delta	3.1.0	3.0.0	3.0.0
Flink	1.18.1	1.18.1	1.18.0
HBase	2.4.17	2.4.17	2.4.17
HCatalog	3.1.3	3.1.3	3.1.3
Hadoop	3.3.6	3.3.6	3.3.6

	emr-7.2.0	emr-7.1.0	emr-7.0.0
Hive	3.1.3	3.1.3	3.1.3
Hudi	0.14.1-amzn-1	0.14.1-amzn-0	0.14.0-amzn-1
Hue	4.11.0	4.11.0	4.11.0
Iceberg	1.5.0-amzn-0	1.4.3-amzn-0	1.4.2-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0	1.5.0
Livy	0.8.0	0.8.0	0.7.1
MXNet	1.9.1	1.9.1	1.9.1
Oozie	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3	5.1.3
Pig	0.17.0	0.17.0	0.17.0
Presto	0.285	0.284	0.283
Spark	3.5.1	3.5.0	3.5.0
Sqoop	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0
Tez	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	436	435	426
Zeppelin	0.10.1	0.10.1	0.10.1
ZooKeeper	3.9.1	3.9.1	3.5.10

7.2.0 release notes

The following release notes include information for Amazon EMR release 7.2.0. Changes are relative to 7.2.0.

New features

- **Application upgrades –** Amazon EMR 7.2.0 application upgrades include Iceberg 1.5.0-amzn-0 and Delta 3.1.0.
- Amazon EMR adds support so that you can use other applications such as HBase, Flink, and Hive with the Amazon S3 Express One Zone storage class.
- This release adds the capability to read restored objects, so you can read Glacier objects from an S3 location with the `S3A` protocol. This capability works with Spark, Flink, and Hive.
- Node labels with managed scaling - Use node labels with your instances based on market type or node type to improve automatic scaling from Amazon EMR. For more information, see Using managed scaling in Amazon EMR.

Known issues

- Python 3.11 isn't supported with EMR Studio.

Changes, enhancements, and resolved issues

- This release fixes a deadlock issue that can occur during internal step cleanup operations. This operation manages the life cycle of steps as they complete on the EMR cluster. This issue affects critical Amazon EMR operations, such as step operation and scaling.
- This release resolves an issue where custom clusters with custom AMIs that have certain pre-existing log files can cause the Amazon EMR log management daemon to fail.
- Amazon EMR 7.2.0 upgrades the Amazon EMR daemon responsible for cluster management and monitoring activities from AWS SDK v1 to v2.

- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2023.5.20240819.0	6.1.102-111.182.amzn2023	August 20, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne),

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
			Israel (Tel Aviv), Canada West (Calgary)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2023.5.2 240730.	6.1.97-104.177.amz n2023	August 2, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Eu rope (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing) , China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2023.5.20240722.	6.1.97-104.177.amzn2023	July 24, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2023.5.2 240708.	6.1.96-102.177.amz n2023	July 8th, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), A sia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), Canada West (Calgary), AWS GovCloud (US-West), AWS GovCloud (US-East),China (Beijing), China (Ningxia)

7.2.0 default Java versions

Amazon EMR release 7.2 and higher ships with Amazon Corretto 17 (built on OpenJDK) by default for applications that support Corretto 17 (JDK 17), with the exception of Apache Livy.

The following table shows the default Java versions for applications in Amazon EMR 7.2.0. If you want to change the default JVM on your cluster, follow the instructions in Configure applications to use a specific Java Virtual Machine for each application that runs on the cluster. You can only use one Java runtime version for a cluster. Amazon EMR doesn't support running different nodes or applications on different runtime versions on the same cluster.

Application	Java / Amazon Corretto version (default is bold)
Delta	**17**, 11, 8
Flink	**17**, 11, 8
HBase	11, **8**
HCatalog	**17**, 11, 8
Hadoop	**17**, 11, 8
Hive	**17**, 11, 8
Hudi	**17**, 11, 8
Iceberg	**17**, 11, 8
Livy	17, 11, **8**
Oozie	**17**, 11, 8
Phoenix	**8**
PrestoDB	**8**
Spark	**17**, 11, 8
Spark RAPIDS	**17**, 11, 8

Application	Java / Amazon Corretto version (default is bold)
Sqoop	**8**
Tez	**17**, 11, 8
Trino	**17**
Zeppelin	**8**
Pig	**8**
Zookeeper	**8**

7.2.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
adot-java-agent	1.31.0	A Java Agent that collects metrics from application daemons.
delta	3.1.0	Delta lake is an open table format for huge analytic datasets

Component	Version	Description
emr-amazon-cloudwatch-agent	1.300032.2-amzn-0	An application that collects internal system-level metrics and custom application metrics from Amazon EC2 instances.
emr-ddb	5.4.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.11.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.15.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.16.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.32.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.11.0	EMR S3Select Connector
emr-wal-cli	1.2.2	Cli used for emrwal list/dele tion.
emrfs	2.63.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.18.1-amzn-1	Apache Flink command line client scripts and applications.

Component	Version	Description
flink-jobmanager-config	1.18.1-amzn-1	Managing resources on EMR nodes for Apache Flink JobManager.
hadoop-client	3.3.6-amzn-4	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.3.6-amzn-4	HDFS node-level service for storing blocks.
hadoop-hdfs-library	3.3.6-amzn-4	HDFS command-line client and library
hadoop-hdfs-namenode	3.3.6-amzn-4	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	3.3.6-amzn-4	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	3.3.6-amzn-4	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.3.6-amzn-4	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.3.6-amzn-4	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.3.6-amzn-4	YARN service for managing containers on an individual node.

Component	Version	Description
hadoop-yarn-resour cemanager	3.3.6-amzn-4	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.3.6-amzn-4	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	2.4.17-amzn-6	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.4.17-amzn-6	Service for serving one or more HBase regions.
hbase-client	2.4.17-amzn-6	HBase command-line client.
hbase-rest-server	2.4.17-amzn-6	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	2.4.17-amzn-6	Service providing a Thrift endpoint to HBase.
hbase-operator-tools	2.4.17-amzn-6	Repair tool for Apache HBase clusters.
hcatalog-client	3.1.3-amzn-11	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.3-amzn-11	Service providing HCatalog, a table and storage management layer for distributed applications.

Component	Version	Description
hcatalog-webhcat-server	3.1.3-amzn-11	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.3-amzn-11	Hive command line client.
hive-hbase	3.1.3-amzn-11	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-11	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	3.1.3-amzn-11	Service for accepting Hive queries as web requests.
hudi	0.14.1-amzn-1	Incremental processing framework to power data pipeline at low latency and high efficiency.
hudi-presto	0.14.1-amzn-1	Bundle library for running Presto with Hudi.
hudi-trino	0.14.1-amzn-1	Bundle library for running Trino with Hudi.
hudi-spark	0.14.1-amzn-1	Bundle library for running Spark with Hudi.
hue-server	4.11.0	Web application for analyzing data using Hadoop ecosystem applications
iceberg	1.5.0-amzn-0	Apache Iceberg is an open table format for huge analytic datasets

Component	Version	Description
jupyterhub	1.5.0	Multi-user server for Jupyter notebooks
livy-server	0.8.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.9.1	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68+	MariaDB database server.
nvidia-cuda	11.8.0	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.7.0	Open Source Computer Vision Library.
phoenix-library	5.1.3	The phoenix libraries for server and client
phoenix-connectors	5.1.3	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	6.0.0	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API

Component	Version	Description
presto-coordinator	0.285-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.285-amzn-0	Service for executing pieces of a query.
presto-client	0.285-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
trino-coordinator	436-amzn-0	Service for accepting queries and managing query execution among trino-wor kers.
trino-worker	436-amzn-0	Service for executing pieces of a query.
trino-client	436-amzn-0	Trino command-line client which is installed on an HA cluster's stand-by masters where Trino server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.3.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.5.1-amzn-0	Spark command-line clients.

Component	Version	Description
spark-history-server	3.5.1-amzn-0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	3.5.1-amzn-0	In-memory execution engine for YARN.
spark-yarn-slave	3.5.1-amzn-0	Apache Spark libraries needed by YARN slaves.
spark-rapids	24.02.0-amzn-0	Nvidia Spark RAPIDS plugin that accelerates Apache Spark with GPUs.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.11.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.10.2-amzn-9	The tez YARN application and libraries.
tez-on-worker	0.10.2-amzn-9	The tez YARN application and libraries for worker nodes.
zeppelin-server	0.10.1	Web-based notebook that enables interactive data analytics.

Component	Version	Description
zookeeper-server	3.9.1-amzn-1	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.9.1-amzn-1	ZooKeeper command line client.

7.2.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-7.2.0 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode,

Classifications	Description	Reconfiguration Actions
		SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.

Classifications	Description	Reconfiguration Actions
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-ses sion.properties settings for Kubernetes/Yarn session.	Restarts Flink history server.
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Restarts Flink history server.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts PhoenixQu eryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	This classification should not be reconfigured.

Classifications	Description	Reconfiguration Actions
hdfs-env	Change values in the HDFS environment.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat server.
hive	Amazon EMR-curated settings for Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.
hive-env	Change values in the Hive environment.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Not available.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Classifications	Description	Reconfiguration Actions
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Change Hadoop's KMS java home	Not available.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Change values in the Hudi environment.	Not available.
hudi-defaults	Change values in Hudi's hudi-defaults.conf file.	Not available.
iceberg-defaults	Change values in Iceberg's iceberg-defaults.conf file.	Not available.
delta-defaults	Change values in Delta's delta-defaults.conf file.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.

Classifications	Description	Reconfiguration Actions
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j2	Change Livy log4j2.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.

Classifications	Description	Reconfiguration Actions
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hb ase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServe r.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.pro perties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoDB)
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticat or	Change values in Presto's password-authenticator.prop erties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoDB)

Classifications	Description	Reconfiguration Actions
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-lakeformat ion	Change values in Presto's lakeformation.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
trino-log	Change values in Trino's log.properties file.	Restarts Trino-Server (for Trino)
trino-config	Change values in Trino's config.properties file.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Change values in Trino's password-authenticator.prop erties file.	Restarts Trino-Server (for Trino)
trino-env	Change values in Trino's trino-env.sh file.	Restarts Trino-Server (for Trino)
trino-node	Change values in Trino's node.properties file.	Not available.
trino-connector-blackhole	Change values in Trino's blackhole.properties file.	Not available.
trino-connector-cassandra	Change values in Trino's cassandra.properties file.	Not available.
trino-connector-delta	Change values in Trino's delta.properties file.	Restarts Trino-Server (for Trino)

Classifications	Description	Reconfiguration Actions
trino-connector-hive	Change values in Trino's hive.properties file.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Change values in Trino's exchange-manager.properties file.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Change values in Trino's iceberg.properties file.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Change values in Trino's hudi.properties file.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Change values in Trino's jmx.properties file.	Not available.
trino-connector-kafka	Change values in Trino's kafka.properties file.	Not available.
trino-connector-localfile	Change values in Trino's localfile.properties file.	Not available.
trino-connector-memory	Change values in Trino's memory.properties file.	Not available.
trino-connector-mongodb	Change values in Trino's mongodb.properties file.	Not available.
trino-connector-mysql	Change values in Trino's mysql.properties file.	Not available.
trino-connector-postgresql	Change values in Trino's postgresql.properties file.	Not available.
trino-connector-raptor	Change values in Trino's raptor.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
trino-connector-redis	Change values in Trino's redis.properties file.	Not available.
trino-connector-redshift	Change values in Trino's redshift.properties file.	Not available.
trino-connector-tpch	Change values in Trino's tpch.properties file.	Not available.
trino-connector-tpcds	Change values in Trino's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-logback	Change values in kms-logba ck.xml file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.

Classifications	Description	Reconfiguration Actions
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j2	Change values in Spark's log4j2.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restart Oozie and HiveServer2.
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-logback	Change values in ZooKeeper's logback.xml file.	Restarts Zookeeper server.
emr-metrics	Change emr metric settings for this node.	Restarts the CloudWatc hAgent service.

7.2.0 change log

Change log for 7.2.0 release and release notes

Date	Event	Description
2024-07-25	Docs publication	Amazon EMR 7.2.0 release notes first published
2024-07-16	Initial release	Amazon EMR 7.2.0 first deployed to initial commercia l Regions

Amazon EMR release 7.1.0

7.1.0 application versions

The following applications are supported in this release: AmazonCloudWatchAgent, Delta, Flink, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, Iceberg, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Trino, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-7.2.0	**emr-7.1.0**	**emr-7.0.0**
AWS SDK for Java	2.23.18, 1.12.705	2.23.18, 1.12.656	2.20.160-amzn-0, 1.12.569
Python	3.9	3.9	3.9
Scala	2.12.17	2.12.17	2.12.17
AmazonCloudWatchAgent	1.300032.2	1.300032.2	1.300031.1
Delta	3.1.0	3.0.0	3.0.0
Flink	1.18.1	1.18.1	1.18.0
HBase	2.4.17	2.4.17	2.4.17
HCatalog	3.1.3	3.1.3	3.1.3
Hadoop	3.3.6	3.3.6	3.3.6
Hive	3.1.3	3.1.3	3.1.3
Hudi	0.14.1-amzn-1	0.14.1-amzn-0	0.14.0-amzn-1
Hue	4.11.0	4.11.0	4.11.0

	emr-7.2.0	emr-7.1.0	emr-7.0.0
Iceberg	1.5.0-amzn-0	1.4.3-amzn-0	1.4.2-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0	1.5.0
Livy	0.8.0	0.8.0	0.7.1
MXNet	1.9.1	1.9.1	1.9.1
Mahout	-	-	-
Oozie	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3	5.1.3
Pig	0.17.0	0.17.0	0.17.0
Presto	0.285	0.284	0.283
Spark	3.5.1	3.5.0	3.5.0
Sqoop	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0
Tez	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	436	435	426
Zeppelin	0.10.1	0.10.1	0.10.1
ZooKeeper	3.9.1	3.9.1	3.5.10

7.1.0 release notes

The following release notes include information for Amazon EMR release 7.1.0. Changes are relative to 7.1.0.

New features

- **Application upgrades** – Amazon EMR 7.1.0 application upgrades include Livy 0.8.0, Trino 435, and ZooKeeper 3.9.1.
- **Unhealthy node replacement** – With Amazon EMR 7.1.0 and higher, unhealthy node replacement is enabled by default, so Amazon EMR will gracefully replace your unhealthy nodes. To avoid affecting your existing workflows on Amazon EMR releases 7.0.0 and lower, unhealthy node replacement is disabled if you enabled termination protection in your cluster.
- **CloudWatch Agent** – Configure the CloudWatch agent to use additional system metrics, add application metrics, and change metrics destination with the Amazon EMR configuration API.

Known issues

- Python 3.11 isn't supported with EMR Studio.

Changes, enhancements, and resolved issues

- While Amazon EMR 7.1.0 supports Python 3.9 by default, Livy 0.8.0 and Spark in Amazon EMR 7.1.0 support Python 3.11.
- This release fixes the issue of needing to run each line one at a time when using PySpark with Python version 3.11.
- **Zeppelin upgrade** – Amazon EMR 7.1.0 includes an upgrade of Zeppelin to the AWS SDK for Java v2. This upgrade enables a Zeppelin S3 Notebook to accept a custom encryption materials provider. The AWS SDK for Java v2 removes the EncryptionMaterialsProvider interface. When you upgrade to Amazon EMR 7.1.0, you must implement the Keyring interface if you want to use custom encryption. For an example of how to implement the Keyring interface, see KmsKeyring.java.
- When upgrading to Amazon EMR release 7.1.0, change your custom key provider for local disk encryption to generate keys using the AES algorithm of AES/GCM/NoPadding. If you don't update the algorithm, cluster creation might fail with the error `Local disk encryption failed on master instance (i-123456789) due to internal error.` For more information about creating a custom key provider, see Creating a custom key provider.
- Amazon EMR 7.1.0 improves the resiliency of a node under low disk space conditions by improving log truncation logic for files with open file handles.

- This release enhances the encoding and decoding logic to minimize risk of data corruption and node failure with Amazon EMR daemons read and write files when restarting a node.

- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2023.5.2 240819.	6.1.102-111.182.amzn2023	August 20, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne),

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
			Israel (Tel Aviv), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2023.5.2 240730.	6.1.97-104.177.amzn2023	August 2, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsRelease Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2023.5.2 240722.	6.1.97-104.177.amzn2023	July 24, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2023.5.2 240708.	6.1.96-102.177.amzn2023	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2023.3.2 240219.	6.1.77-99.164.amzn2023	May 8th, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), Canada West (Calgary), AWS GovCloud (US-West), AWS GovCloud (US-East),China (Beijing), China (Ningxia)

7.1.0 default Java versions

Amazon EMR release 7.1 and higher ships with Amazon Corretto 17 (built on OpenJDK) by default for applications that support Corretto 17 (JDK 17), with the exception of Apache Livy.

The following table shows the default Java versions for applications in Amazon EMR 7.1.0. If you want to change the default JVM on your cluster, follow the instructions in Configure applications to use a specific Java Virtual Machine for each application that runs on the cluster. You can only use one Java runtime version for a cluster. Amazon EMR doesn't support running different nodes or applications on different runtime versions on the same cluster.

Application	Java / Amazon Corretto version (default is bold)
Delta	**17**, 11, 8
Flink	**17**, 11, 8
HBase	11, **8**
HCatalog	**17**, 11, 8
Hadoop	**17**, 11, 8
Hive	**17**, 11, 8
Hudi	**17**, 11, 8
Iceberg	**17**, 11, 8
Livy	17, 11, **8**
Oozie	**17**, 11, 8
Phoenix	**8**
PrestoDB	**8**
Spark	**17**, 11, 8
Spark RAPIDS	**17**, 11, 8

Application	Java / Amazon Corretto version (default is bold)
Sqoop	**8**
Tez	**17**, 11, 8
Trino	**17**
Zeppelin	**8**
Pig	**8**
Zookeeper	**8**

7.1.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
adot-java-agent	1.31.0	A Java Agent that collects metrics from application daemons.
delta	3.0.0	Delta lake is an open table format for huge analytic datasets

Component	Version	Description
emr-amazon-cloudwatch-agent	1.300032.2-amzn-0	An application that collects internal system-level metrics and custom application metrics from Amazon EC2 instances.
emr-ddb	5.3.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.10.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.14.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.7.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.31.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.10.0	EMR S3Select Connector
emr-wal-cli	1.2.2	Cli used for emrwal list/dele tion.
emrfs	2.62.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.18.1-amzn-0	Apache Flink command line client scripts and applications.

Component	Version	Description
flink-jobmanager-config	1.18.1-amzn-0	Managing resources on EMR nodes for Apache Flink JobManager.
hadoop-client	3.3.6-amzn-3	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.3.6-amzn-3	HDFS node-level service for storing blocks.
hadoop-hdfs-library	3.3.6-amzn-3	HDFS command-line client and library
hadoop-hdfs-namenode	3.3.6-amzn-3	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	3.3.6-amzn-3	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	3.3.6-amzn-3	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.3.6-amzn-3	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.3.6-amzn-3	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.3.6-amzn-3	YARN service for managing containers on an individual node.

Component	Version	Description
hadoop-yarn-resourcemanager	3.3.6-amzn-3	YARN service for allocating and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.3.6-amzn-3	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	2.4.17-amzn-5	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.4.17-amzn-5	Service for serving one or more HBase regions.
hbase-client	2.4.17-amzn-5	HBase command-line client.
hbase-rest-server	2.4.17-amzn-5	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	2.4.17-amzn-5	Service providing a Thrift endpoint to HBase.
hbase-operator-tools	2.4.17-amzn-5	Repair tool for Apache HBase clusters.
hcatalog-client	3.1.3-amzn-10	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.3-amzn-10	Service providing HCatalog, a table and storage management layer for distributed applications.

Component	Version	Description
hcatalog-webhcat-server	3.1.3-amzn-10	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.3-amzn-10	Hive command line client.
hive-hbase	3.1.3-amzn-10	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-10	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	3.1.3-amzn-10	Service for accepting Hive queries as web requests.
hudi	0.14.1-amzn-0	Incremental processing framework to power data pipeline at low latency and high efficiency.
hudi-presto	0.14.1-amzn-0	Bundle library for running Presto with Hudi.
hudi-trino	0.14.1-amzn-0	Bundle library for running Trino with Hudi.
hudi-spark	0.14.1-amzn-0	Bundle library for running Spark with Hudi.
hue-server	4.11.0	Web application for analyzing data using Hadoop ecosystem applications
iceberg	1.4.3-amzn-0	Apache Iceberg is an open table format for huge analytic datasets

Component	Version	Description
jupyterhub	1.5.0	Multi-user server for Jupyter notebooks
livy-server	0.8.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.9.1	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68+	MariaDB database server.
nvidia-cuda	11.8.0	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.7.0	Open Source Computer Vision Library.
phoenix-library	5.1.3	The phoenix libraries for server and client
phoenix-connectors	5.1.3	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.3	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API

Component	Version	Description
presto-coordinator	0.284-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.284-amzn-0	Service for executing pieces of a query.
presto-client	0.284-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
trino-coordinator	435-amzn-0	Service for accepting queries and managing query execution among trino-wor kers.
trino-worker	435-amzn-0	Service for executing pieces of a query.
trino-client	435-amzn-0	Trino command-line client which is installed on an HA cluster's stand-by masters where Trino server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.3.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.5.0-amzn-1	Spark command-line clients.

Component	Version	Description
spark-history-server	3.5.0-amzn-1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	3.5.0-amzn-1	In-memory execution engine for YARN.
spark-yarn-slave	3.5.0-amzn-1	Apache Spark libraries needed by YARN slaves.
spark-rapids	23.10.0-amzn-1	Nvidia Spark RAPIDS plugin that accelerates Apache Spark with GPUs.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.11.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.10.2-amzn-8	The tez YARN application and libraries.
tez-on-worker	0.10.2-amzn-8	The tez YARN application and libraries for worker nodes.
zeppelin-server	0.10.1	Web-based notebook that enables interactive data analytics.

Component	Version	Description
zookeeper-server	3.9.1-amzn-0	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.9.1-amzn-0	ZooKeeper command line client.

7.1.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-7.1.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
docker-conf	Change docker related settings.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.

Classifications	Description
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-session	Change Flink log4j-session.properties settings for Kubernetes/Yarn session.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-env	Change values in the HDFS environment.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.

Classifications	Description
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive	Amazon EMR-curated settings for Apache Hive.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file

Classifications	Description
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-java-home	Change Hadoop's KMS java home
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
hudi-env	Change values in the Hudi environment.
hudi-defaults	Change values in Hudi's hudi-defaults.conf file.
iceberg-defaults	Change values in Iceberg's iceberg-defaults.c onf file.
delta-defaults	Change values in Delta's delta-defaults.conf file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.

Classifications	Description
livy-env	Change values in the Livy environment.
livy-log4j2	Change Livy log4j2.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.

Classifications	Description
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-lakeformation	Change values in Presto's lakeformation.prop erties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.

Classifications	Description
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.
trino-log	Change values in Trino's log.properties file.
trino-config	Change values in Trino's config.properties file.
trino-password-authenticator	Change values in Trino's password-authenticator.properties file.
trino-env	Change values in Trino's trino-env.sh file.
trino-node	Change values in Trino's node.properties file.
trino-connector-blackhole	Change values in Trino's blackhole.properties file.
trino-connector-cassandra	Change values in Trino's cassandra.properties file.
trino-connector-delta	Change values in Trino's delta.properties file.
trino-connector-hive	Change values in Trino's hive.properties file.
trino-exchange-manager	Change values in Trino's exchange-manager.properties file.
trino-connector-iceberg	Change values in Trino's iceberg.properties file.
trino-connector-hudi	Change values in Trino's hudi.properties file.
trino-connector-jmx	Change values in Trino's jmx.properties file.

Classifications	Description
trino-connector-kafka	Change values in Trino's kafka.properties file.
trino-connector-localfile	Change values in Trino's localfile.properties file.
trino-connector-memory	Change values in Trino's memory.properties file.
trino-connector-mongodb	Change values in Trino's mongodb.properties file.
trino-connector-mysql	Change values in Trino's mysql.properties file.
trino-connector-postgresql	Change values in Trino's postgresql.properties file.
trino-connector-raptor	Change values in Trino's raptor.properties file.
trino-connector-redis	Change values in Trino's redis.properties file.
trino-connector-redshift	Change values in Trino's redshift.properties file.
trino-connector-tpch	Change values in Trino's tpch.properties file.
trino-connector-tpcds	Change values in Trino's tpcds.properties file.
ranger-kms-dbks-site	Change values in dbks-site.xml file of Ranger KMS.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.
ranger-kms-env	Change values in the Ranger KMS environme nt.
ranger-kms-logback	Change values in kms-logback.xml file of Ranger KMS.

Classifications	Description
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j2	Change values in Spark's log4j2.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zeppelin-site	Change configuration settings in zeppelin-site.xml.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-logback	Change values in ZooKeeper's logback.xml file.

Classifications	Description
emr-metrics	Change emr metric settings for this node.

7.1.0 change log

Change log for 7.1.0 release and release notes

Date	Event	Description
2024-05-10	Docs publication	Amazon EMR 7.1.0 release notes first published
2024-05-08	Deployment complete	Amazon EMR 7.1.0 fully deployed to all supported Regions
2024-04-23	Initial release	Amazon EMR 7.1.0 first deployed to initial commercia l Regions

Amazon EMR release 7.0.0

7.0.0 application versions

The following applications are supported in this release: AmazonCloudWatchAgent, Delta, Flink, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, Iceberg, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Trino, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases

- Application versions in Amazon EMR 4.x releases

Application version information

	emr-7.2.0	emr-7.1.0	emr-7.0.0
AWS SDK for Java	2.23.18, 1.12.705	2.23.18, 1.12.656	2.20.160-amzn-0, 1.12.569
Python	3.9	3.9	3.9
Scala	2.12.17	2.12.17	2.12.17
AmazonCloudWatchAgent	1.300032.2	1.300032.2	1.300031.1
Delta	3.1.0	3.0.0	3.0.0
Flink	1.18.1	1.18.1	1.18.0
HBase	2.4.17	2.4.17	2.4.17
HCatalog	3.1.3	3.1.3	3.1.3
Hadoop	3.3.6	3.3.6	3.3.6
Hive	3.1.3	3.1.3	3.1.3
Hudi	0.14.1-amzn-1	0.14.1-amzn-0	0.14.0-amzn-1
Hue	4.11.0	4.11.0	4.11.0
Iceberg	1.5.0-amzn-0	1.4.3-amzn-0	1.4.2-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0	1.5.0
Livy	0.8.0	0.8.0	0.7.1

	emr-7.2.0	emr-7.1.0	emr-7.0.0
MXNet	1.9.1	1.9.1	1.9.1
Mahout	-	-	-
Oozie	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3	5.1.3
Pig	0.17.0	0.17.0	0.17.0
Presto	0.285	0.284	0.283
Spark	3.5.1	3.5.0	3.5.0
Sqoop	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0
Tez	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	436	435	426
Zeppelin	0.10.1	0.10.1	0.10.1
ZooKeeper	3.9.1	3.9.1	3.5.10

7.0.0 release notes

The following release notes include information for Amazon EMR release 7.0.0. Changes are relative to 6.15.0.

New features

- **Application upgrades** – Amazon EMR 7.0.0 application upgrades include Python 3.9, Spark 3.5, Flink 1.18, and Delta 3.0. This release also adds support for the Amazon CloudWatch agent application and removes support for Ganglia.
- **Amazon Corretto 17** – Amazon EMR releases 7.0 and higher ship with Amazon Corretto 17 (built on OpenJDK) by default for applications that support Corretto 17 (JDK 17), with the exception

of Apache Livy. For more information about the supported JDK versions for applications in this release, see 7.0.0 default Java versions.

- **Amazon Linux 2023 –** With the 7.0 release, Amazon EMR clusters now run on AL2023 by default. For information on how this affects the default AMI version, see Software update considerations in the *Amazon EMR Management Guide*. Also note that AL2023 removed Python 2.7, so any components that require Python should now be written with Python 3.
- **S3 on Outposts with `s3a` –** Amazon EMR now supports Amazon S3 on Outposts buckets with the `s3a` file system. For more information about S3 on Outposts, see What is S3 on Outposts? in the *Amazon Simple Storage Service User Guide*.

Known issues

- Note that you can't run more than one line at a time if you are using PySpark with Python versions 3.10 or later. You must run each line one at a time.

Changes, enhancements, and resolved issues

- To maintain the state of all the instances in a cluster, Amazon EMR integrates with Apache YARN, Apache HDFS, and Kerberos. With 7.0, we enhanced these integrations to reliably remove the state of instances that are terminated due to scaling or other operations. This is particularly important for long-lived clusters that have managed scaling enabled, as they scale frequently and accumulate thousands of terminated instances over their lifetimes.
- This release improves the Kerberos configuration to only include support for AES-based ciphers. Kerberos KDC with non-AES based ciphers are no longer supported with EMR clusters that run on Amazon EMR releases 7.0.0 and higher. AES-based ciphers offer the strongest security for your clusters.
- As part of the AWS SDK 2.x migration, Amazon EMR 7.0 includes an update to the Spark Kinesis connector for compatibility. This update isn't available in the community version of Apache Spark. If you use the Spark Kinesis connector from an Amazon EMR release lower than 7.0, you must migrate your application codes to run on SDK 2.x before you can migrate your workloads to Amazon EMR 7.0. For more information, see Migrating Spark Kinesis connector to SDK 2.x for Amazon EMR 7.0.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2

release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2023.5.2 240819.	6.1.102-111.182.amzn2023	August 20, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne),

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
			Israel (Tel Aviv), Canada West (Calgary)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2023.5.2 240730.	6.1.97-104.177.amzn2023	August 2, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2023.5.20240722.	6.1.97-104.177.amzn2023	July 24, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2023.5.2 240708.	6.1.96-102.177.amz n2023	July 23, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), Canada West (Calgary), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2023.3.2 240304.	6.1.79-99.164.amzn2023	March 12, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), Canada West (Calgary), AWS GovCloud (US-West), AWS GovCloud (US-East),China (Beijing), China (Ningxia)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2023.3.2 240219.	6.1.77-99.164.amzn2023	March 1, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), Canada West (Calgary), AWS GovCloud (US-West), AWS GovCloud (US-East),China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2023.3.2 240205.	6.1.75-99.163.amzn2023	February 19, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), Canada West (Calgary), AWS GovCloud (US-West), AWS GovCloud (US-East),China (Beijing), China (Ningxia)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2023.3.2 240122.	6.1.72-96.166.amzn2023	February 5, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), Canada West (Calgary), AWS GovCloud (US-West), AWS GovCloud (US-East),China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2023.3.2 240108.	6.1.72-96.166.amzn2023	January 24, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), Canada West (Calgary), AWS GovCloud (US-West), AWS GovCloud (US-East),China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2023.3.20231211.	6.1.66-91.160.amzn2023	December 19, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

7.0.0 default Java versions

Amazon EMR release 7.0 and higher ships with Amazon Corretto 17 (built on OpenJDK) by default for applications that support Corretto 17 (JDK 17), with the exception of Apache Livy.

The following table shows the default Java versions for applications in Amazon EMR 7.0.0. If you want to change the default JVM on your cluster, follow the instructions in Configure applications to use a specific Java Virtual Machine for each application that runs on the cluster. You can only use one Java runtime version for a cluster. Amazon EMR doesn't support running different nodes or applications on different runtime versions on the same cluster.

Application	Java / Amazon Corretto version (default is bold)
Delta	**17**, 11, 8
Flink	11, **8**
HBase	11, **8**
HCatalog	**17**, 11, 8
Hadoop	**17**, 11, 8
Hive	**17**, 11, 8
Hudi	**17**, 11, 8
Iceberg	**17**, 11, 8
Livy	17, 11, **8**
Oozie	**17**, 11, 8
Phoenix	**8**
PrestoDB	**8**
Spark	**17**, 11, 8
Spark RAPIDS	**17**, 11, 8

Application	Java / Amazon Corretto version (default is bold)
Sqoop	**8**
Tez	**17**, 11, 8
Trino	**17**
Zeppelin	**8**
Pig	**8**
Zookeeper	**8**

7.0.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
adot-java-agent	1.31.0	A Java Agent that collects metrics from application daemons.
delta	3.0.0	Delta lake is an open table format for huge analytic datasets

Component	Version	Description
delta-standalone-connectors	0.6.0	Delta Connectors provide different runtimes to integrate Delta Lake with engines like Flink, Hive and Presto.
emr-amazon-cloudwatch-agent	1.300031.1-amzn-0	An application that collects internal system-level metrics and custom application metrics from Amazon EC2 instances.
emr-ddb	5.2.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.9.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.13.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.7.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.30.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.9.0	EMR S3Select Connector
emr-wal-cli	1.2.1	Cli used for emrwal list/dele tion.

Component	Version	Description
emrfs	2.61.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.18.0-amzn-0	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.18.0-amzn-0	Managing resources on EMR nodes for Apache Flink JobManager.
hadoop-client	3.3.6-amzn-2	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.3.6-amzn-2	HDFS node-level service for storing blocks.
hadoop-hdfs-library	3.3.6-amzn-2	HDFS command-line client and library
hadoop-hdfs-namenode	3.3.6-amzn-2	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	3.3.6-amzn-2	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	3.3.6-amzn-2	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.3.6-amzn-2	Cryptographic key management server based on Hadoop's KeyProvider API.

Component	Version	Description
hadoop-mapred	3.3.6-amzn-2	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.3.6-amzn-2	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.3.6-amzn-2	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.3.6-amzn-2	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	2.4.17-amzn-4	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.4.17-amzn-4	Service for serving one or more HBase regions.
hbase-client	2.4.17-amzn-4	HBase command-line client.
hbase-rest-server	2.4.17-amzn-4	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	2.4.17-amzn-4	Service providing a Thrift endpoint to HBase.
hbase-operator-tools	2.4.17-amzn-4	Repair tool for Apache HBase clusters.

Component	Version	Description
hcatalog-client	3.1.3-amzn-9	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.3-amzn-9	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.3-amzn-9	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.3-amzn-9	Hive command line client.
hive-hbase	3.1.3-amzn-9	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-9	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	3.1.3-amzn-9	Service for accepting Hive queries as web requests.
hudi	0.14.0-amzn-1	Incremental processing framework to power data pipeline at low latency and high efficiency.
hudi-presto	0.14.0-amzn-1	Bundle library for running Presto with Hudi.
hudi-trino	0.14.0-amzn-1	Bundle library for running Trino with Hudi.

Component	Version	Description
hudi-spark	0.14.0-amzn-1	Bundle library for running Spark with Hudi.
hue-server	4.11.0	Web application for analyzing data using Hadoop ecosystem applications
iceberg	1.4.2-amzn-0	Apache Iceberg is an open table format for huge analytic datasets
jupyterhub	1.5.0	Multi-user server for Jupyter notebooks
livy-server	0.7.1-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.9.1	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68+	MariaDB database server.
nvidia-cuda	11.8.0	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.7.0	Open Source Computer Vision Library.

Component	Version	Description
phoenix-library	5.1.3	The phoenix libraries for server and client
phoenix-connectors	5.1.3	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.3	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.283-amzn-1	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.283-amzn-1	Service for executing pieces of a query.
presto-client	0.283-amzn-1	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
trino-coordinator	426-amzn-1	Service for accepting queries and managing query execution among trino-wor kers.
trino-worker	426-amzn-1	Service for executing pieces of a query.

Component	Version	Description
trino-client	426-amzn-1	Trino command-line client which is installed on an HA cluster's stand-by masters where Trino server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.1.3	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.5.0-amzn-0	Spark command-line clients.
spark-history-server	3.5.0-amzn-0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	3.5.0-amzn-0	In-memory execution engine for YARN.
spark-yarn-slave	3.5.0-amzn-0	Apache Spark libraries needed by YARN slaves.
spark-rapids	23.10.0-amzn-0	Nvidia Spark RAPIDS plugin that accelerates Apache Spark with GPUs.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.11.0	TensorFlow open source software library for high performance numerical computation.

Component	Version	Description
tez-on-yarn	0.10.2-amzn-7	The tez YARN application and libraries.
tez-on-worker	0.10.2-amzn-7	The tez YARN application and libraries for worker nodes.
webserver	2.4.58	Apache HTTP server.
zeppelin-server	0.10.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.5.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.5.10	ZooKeeper command line client.

7.0.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see [Configure applications](#).

emr-7.0.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.

Classifications	Description
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
docker-conf	Change docker related settings.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-session	Change Flink log4j-session.properties settings for Kubernetes/Yarn session.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.

Classifications	Description
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-env	Change values in the HDFS environment.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive	Amazon EMR-curated settings for Apache Hive.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.

Classifications	Description
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-java-home	Change Hadoop's KMS java home
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
hudi-env	Change values in the Hudi environment.
hudi-defaults	Change values in Hudi's hudi-defaults.conf file.
iceberg-defaults	Change values in Iceberg's iceberg-defaults.c onf file.
delta-defaults	Change values in Delta's delta-defaults.conf file.

Classifications	Description
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j2	Change Livy log4j2.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.

Classifications	Description
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-lakeformation	Change values in Presto's lakeformation.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.

Classifications	Description
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.
trino-log	Change values in Trino's log.properties file.
trino-config	Change values in Trino's config.properties file.
trino-password-authenticator	Change values in Trino's password-authenticator.properties file.
trino-env	Change values in Trino's trino-env.sh file.
trino-node	Change values in Trino's node.properties file.
trino-connector-blackhole	Change values in Trino's blackhole.properties file.
trino-connector-cassandra	Change values in Trino's cassandra.properties file.
trino-connector-delta	Change values in Trino's delta.properties file.

Classifications	Description
trino-connector-hive	Change values in Trino's hive.properties file.
trino-exchange-manager	Change values in Trino's exchange-manager.p roperties file.
trino-connector-iceberg	Change values in Trino's iceberg.properties file.
trino-connector-hudi	Change values in Trino's hudi.properties file.
trino-connector-jmx	Change values in Trino's jmx.properties file.
trino-connector-kafka	Change values in Trino's kafka.properties file.
trino-connector-localfile	Change values in Trino's localfile.properties file.
trino-connector-memory	Change values in Trino's memory.properties file.
trino-connector-mongodb	Change values in Trino's mongodb.properties file.
trino-connector-mysql	Change values in Trino's mysql.properties file.
trino-connector-postgresql	Change values in Trino's postgresql.properties file.
trino-connector-raptor	Change values in Trino's raptor.properties file.
trino-connector-redis	Change values in Trino's redis.properties file.
trino-connector-redshift	Change values in Trino's redshift.properties file.
trino-connector-tpch	Change values in Trino's tpch.properties file.
trino-connector-tpcds	Change values in Trino's tpcds.properties file.

Classifications	Description
ranger-kms-dbks-site	Change values in dbks-site.xml file of Ranger KMS.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.
ranger-kms-env	Change values in the Ranger KMS environme nt.
ranger-kms-logback	Change values in kms-logback.xml file of Ranger KMS.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j2	Change values in Spark's log4j2.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.

Classifications	Description
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zeppelin-site	Change configuration settings in zeppelin-site.xml.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

7.0.0 change log

Change log for 7.0.0 release and release notes

Date	Event	Description
2023-12-29	Deployment complete	Amazon EMR 7.0.0 fully deployed to all supported Regions
2023-12-28	Docs publication	Amazon EMR 7.0.0 release notes first published
2023-12-19	Initial release	Amazon EMR 7.0.0 first deployed to initial commercial Regions

Amazon EMR 6.x release versions

This section contains application versions, release notes, component versions, and configuration classifications available in each Amazon EMR 6.x release version.

When you launch a cluster, you can choose from multiple releases of Amazon EMR. This allows you to test and use application versions that fit your compatibility requirements. You specify the release number with the *release label*. Release labels are in the form `emr-x.x.x`. For example, `emr-7.2.0`.

New Amazon EMR releases are made available in different Regions over a period of several days, beginning with the first Region on the initial release date. The latest release version may not be available in your Region during this period.

For a comprehensive table of application versions in every Amazon EMR 6.x release, see Application versions in Amazon EMR 6.x releases.

Topics

- Amazon EMR release 6.1.1
- Amazon EMR release 6.1.0
- Amazon EMR release 6.0.1
- Amazon EMR release 6.0.0

Application versions in Amazon EMR 6.x releases

For a comprehensive table that lists the application versions available in each Amazon EMR 6.x release, open Application versions in Amazon EMR 6.x releases in your browser.

Amazon EMR release 6.15.0

6.15.0 application versions

The following applications are supported in this release: Delta, Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, Iceberg, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Trino, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.15.0	emr-6.14.0	emr-6.13.0	emr-6.12.0
AWS SDK for Java	2.20.160-amzn-0, 1.12.569	1.12.543	1.12.513	1.12.490
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7

	emr-6.15.0	emr-6.14.0	emr-6.13.0	emr-6.12.0
Scala	2.12.17	2.12.15	2.12.15	2.12.15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.4.0	2.4.0	2.4.0	2.4.0
Flink	1.17.1	1.17.1	1.17.0	1.17.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.17	2.4.17	2.4.17	2.4.17
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.6	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.14.0-amzn-0	0.13.1-amzn-2	0.13.1-amzn-1	0.13.1-amzn-0
Hue	4.11.0	4.11.0	4.11.0	4.11.0
Iceberg	1.4.0-amzn-0	1.3.1-amzn-0	1.3.0-amzn-1	1.3.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0	1.5.0	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3	5.1.3	5.1.3

	emr-6.15.0	emr-6.14.0	emr-6.13.0	emr-6.12.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.283	0.281	0.281	0.281
Spark	3.4.1	3.4.1	3.4.1	3.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.11.0
Tez	0.10.2	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	426	422	414	414
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

6.15.0 release notes

The following release notes include information for Amazon EMR release 6.15.0. Changes are relative to 6.14.0. For information on the release timeline, see the 6.15.0 change log.

New features

- **Application upgrades** – Amazon EMR 6.15.0 application upgrades include Apache Hadoop 3.3.6, Apache Hudi 0.14.0-amzn-0, Iceberg 1.4.0-amzn-0, and Trino 426.
- **Faster launches for EMR clusters that run on EC2** – It's now up to 35% faster to launch an Amazon EMR on EC2 cluster. With this improvement, most customers can launch their clusters in 5 minutes or less.
- **CodeWhisperer for EMR Studio** – You can now use Amazon CodeWhisperer with Amazon EMR Studio to get real-time recommendations as you write code in JupyterLab. CodeWhisperer can complete your comments, finish single lines of code, make line-by-line recommendations, and generate fully-formed functions.
- **Faster job restart times with Flink** – With Amazon EMR 6.15.0 and higher, several new mechanisms are available for Apache Flink to improve the job restart time during task recovery

or scaling operations. This optimizes the speed of recovery and restart of execution graphs to improve job stability.

- **Table-level and fine-grained access control for open-table formats** – With Amazon EMR 6.15.0 and higher, when you run Spark jobs on Amazon EMR on EC2 clusters that access data in the AWS Glue Data Catalog, you can use AWS Lake Formation to apply table, row, column, and cell level permissions on Hudi, Iceberg, or Delta Lake based tables.
- **Hadoop upgrade** – Amazon EMR 6.15.0 includes an upgrade of Apache Hadoop to version 3.3.6. Hadoop 3.3.6 was the latest version at the time of the Amazon EMR 6.15 deployment, released by Apache in June 2023. Prior releases of Amazon EMR (6.9.0 to 6.14.x) used Hadoop 3.3.3.

  The upgrade includes hundreds of improvements and fixes, and features that include reconfigurable datanode parameters, `DFSAdmin` option to initiate bulk reconfiguration operations on all live datanodes, and a vectored API that allows seek-heavy readers to specify multiple ranges to read. Hadoop 3.3.6 also adds support for HDFS APIs and semantics for its write-ahead log (WAL), so that HBase can run on other storage system implementations. For more information, see the changelogs for versions 3.3.4, 3.3.5, and 3.3.6 in the *Apache Hadoop documentation*.
- **Support for AWS SDK for Java, version 2** - Amazon EMR 6.15.0 applications can use AWS SDK for Java versions 1.12.569 or 2.20.160 if the application supports v2. The AWS SDK for Java 2.x is a major rewrite of the version 1.x code base. It's built on top of Java 8+ and adds several frequently requested features. These include support for non-blocking I/O, and the ability to plug in a different HTTP implementation at runtime. For more information, including a **Migration Guide from SDK for Java v1 to v2**, see the AWS SDK for Java, version 2 guide.

Changes, enhancements, and resolved issues

- To improve your high-availability EMR clusters, this release enables connectivity to Amazon EMR daemons on local host that use IPv6 endpoints.
- This release enables TLS 1.2 for communication with ZooKeeper provisioned on all the primary nodes of your high-availability cluster.
- This release improves the management of ZooKeeper transaction log files that are maintained on primary nodes to minimize scenarios where the log files grow out of bounds and interrupt cluster operations.
- This release makes intra-node communication more resilient for high-availability EMR clusters. This improvement reduces the chance of bootstrap action failures or cluster start failures.

- Tez in Amazon EMR 6.15.0 introduces configurations that you can specify to asynchronously open the input splits in a Tez grouped split. This results in faster performance of read queries when there are a large number of input splits in a single Tez grouped split. For more information, see Tez asynchronous split opening.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240816.0	4.14.350	August 21, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
			(US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)
2.0.20240809.0	4.14.349	August 20, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240719.0	4.14.348	July 25, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 709.1	4.14.348	July 23, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Euro pe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing) , China (Ningxia), Asi a Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 223.0	4.14.336	March 8, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia), Canada West (Calgary)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 131.0	4.14.336	February 14, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240124.0	4.14.336	February 7, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 109.0	4.14.334	January 24, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia), Canada West (Calgary)

OsRelea Label (Amazoı Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 218.0	4.14.330	January 2, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231206.0	4.14.330	December 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 116.0	4.14.328	December 11, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231101.0	4.14.327	November 13, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

6.15.0 default Java versions

Amazon EMR releases 6.12.0 and higher support all applications with Amazon Corretto 8 by default, except for Trino. For Trino, Amazon EMR supports Amazon Corretto 17 by default starting with Amazon EMR release 6.9.0. Amazon EMR also supports some applications with Amazon Corretto 11 and 17. Those applications are listed in the following table. If you want to change the default JVM on your cluster, follow the instructions in Configure applications to use a specific Java Virtual Machine for each application that runs on the cluster. You can only use one Java runtime version for a cluster. Amazon EMR doesn't support running different nodes or applications on different runtime versions on the same cluster.

While Amazon EMR supports both Amazon Corretto 11 and 17 on Apache Spark, Apache Hadoop, and Apache Hive, performance might regress for some workloads when you use these versions of Corretto. We recommend that you test your workloads before you change defaults.

The following table shows the default Java versions for applications in Amazon EMR 6.15.0:

Application	Java / Amazon Corretto version (default is bold)
Delta	17, 11, **8**
Flink	11, **8**
Ganglia	**8**
HBase	11, **8**
HCatalog	17, 11, **8**
Hadoop	17, 11, **8**
Hive	17, 11, **8**
Hudi	17, 11, **8**
Iceberg	17, 11, **8**
Livy	17, 11, **8**
Oozie	17, 11, **8**

Application	Java / Amazon Corretto version (default is bold)
Phoenix	**8**
PrestoDB	**8**
Spark	17, 11, **8**
Spark RAPIDS	17, 11, **8**
Sqoop	**8**
Tez	17, 11, **8**
Trino	**17**
Zeppelin	**8**
Pig	**8**
Zookeeper	**8**

6.15.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.4.0	Delta lake is an open table format for huge analytic datasets
delta-standalone-connectors	0.6.0	Delta Connectors provide different runtimes to integrate Delta Lake with engines like Flink, Hive and Presto.
emr-ddb	5.2.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.8.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.12.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.7.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.29.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.8.0	EMR S3Select Connector
emr-wal-cli	1.2.0	Cli used for emrwal list/dele tion.

Component	Version	Description
emrfs	2.60.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.17.1-amzn-1	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.17.1-amzn-1	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	3.3.6-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.3.6-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	3.3.6-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	3.3.6-amzn-1	HDFS service for tracking file names and block locations.

Component	Version	Description
hadoop-hdfs-journalnode	3.3.6-amzn-1	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	3.3.6-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.3.6-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.3.6-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.3.6-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.3.6-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.3.6-amzn-1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	2.4.17-amzn-3	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.4.17-amzn-3	Service for serving one or more HBase regions.
hbase-client	2.4.17-amzn-3	HBase command-line client.

Component	Version	Description
hbase-rest-server	2.4.17-amzn-3	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	2.4.17-amzn-3	Service providing a Thrift endpoint to HBase.
hbase-operator-tools	2.4.17-amzn-3	Repair tool for Apache HBase clusters.
hcatalog-client	3.1.3-amzn-8	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.3-amzn-8	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.3-amzn-8	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.3-amzn-8	Hive command line client.
hive-hbase	3.1.3-amzn-8	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-8	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	3.1.3-amzn-8	Service for accepting Hive queries as web requests.

Component	Version	Description
hudi	0.14.0-amzn-0	Incremental processing framework to power data pipeline at low latency and high efficiency.
hudi-presto	0.14.0-amzn-0	Bundle library for running Presto with Hudi.
hudi-trino	0.14.0-amzn-0	Bundle library for running Trino with Hudi.
hudi-spark	0.14.0-amzn-0	Bundle library for running Spark with Hudi.
hue-server	4.11.0	Web application for analyzing data using Hadoop ecosystem applications
iceberg	1.4.0-amzn-0	Apache Iceberg is an open table format for huge analytic datasets
jupyterhub	1.5.0	Multi-user server for Jupyter notebooks
livy-server	0.7.1-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.9.1	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68+	MariaDB database server.

Component	Version	Description
nvidia-cuda	11.8.0	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.7.0	Open Source Computer Vision Library.
phoenix-library	5.1.3	The phoenix libraries for server and client
phoenix-connectors	5.1.3	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.3	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.283-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.283-amzn-0	Service for executing pieces of a query.
presto-client	0.283-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.

Component	Version	Description
trino-coordinator	426-amzn-0	Service for accepting queries and managing query execution among trino-wor kers.
trino-worker	426-amzn-0	Service for executing pieces of a query.
trino-client	426-amzn-0	Trino command-line client which is installed on an HA cluster's stand-by masters where Trino server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.4.1-amzn-2	Spark command-line clients.
spark-history-server	3.4.1-amzn-2	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	3.4.1-amzn-2	In-memory execution engine for YARN.
spark-yarn-slave	3.4.1-amzn-2	Apache Spark libraries needed by YARN slaves.
spark-rapids	23.08.1-amzn-0	Nvidia Spark RAPIDS plugin that accelerates Apache Spark with GPUs.

Component	Version	Description
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.11.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.10.2-amzn-6	The tez YARN application and libraries.
tez-on-worker	0.10.2-amzn-6	The tez YARN application and libraries for worker nodes.
webserver	2.4.41+	Apache HTTP server.
zeppelin-server	0.10.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.5.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.5.10	ZooKeeper command line client.

6.15.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-6.15.0 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode,

Classifications	Description	Reconfiguration Actions
		SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-ses sion.properties settings for Kubernetes/Yarn session.	Restarts Flink history server.
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Restarts Flink history server.

Classifications	Description	Reconfiguration Actions
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services SecondaryNamenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.

Classifications	Description	Reconfiguration Actions
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	This classification should not be reconfigured.
hdfs-env	Change values in the HDFS environment.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classifications	Description	Reconfiguration Actions
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat server.
hive	Amazon EMR-curated settings for Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
hive-env	Change values in the Hive environment.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Not available.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.

Classifications	Description	Reconfiguration Actions
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Change Hadoop's KMS java home	Not available.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Change values in the Hudi environment.	Not available.
hudi-defaults	Change values in Hudi's hudi-defaults.conf file.	Not available.
iceberg-defaults	Change values in Iceberg's iceberg-defaults.conf file.	Not available.
delta-defaults	Change values in Delta's delta-defaults.conf file.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.

Classifications	Description	Reconfiguration Actions
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j2	Change Livy log4j2.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hb ase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.

Classifications	Description	Reconfiguration Actions
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServer.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoDB)
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoDB)
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-lakeformation	Change values in Presto's lakeformation.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
trino-log	Change values in Trino's log.properties file.	Restarts Trino-Server (for Trino)
trino-config	Change values in Trino's config.properties file.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Change values in Trino's password-authenticator.prop erties file.	Restarts Trino-Server (for Trino)
trino-env	Change values in Trino's trino-env.sh file.	Restarts Trino-Server (for Trino)
trino-node	Change values in Trino's node.properties file.	Not available.
trino-connector-blackhole	Change values in Trino's blackhole.properties file.	Not available.
trino-connector-cassandra	Change values in Trino's cassandra.properties file.	Not available.
trino-connector-delta	Change values in Trino's delta.properties file.	Restarts Trino-Server (for Trino)
trino-connector-hive	Change values in Trino's hive.properties file.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Change values in Trino's exchange-manager.properties file.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Change values in Trino's iceberg.properties file.	Restarts Trino-Server (for Trino)

Classifications	Description	Reconfiguration Actions
trino-connector-hudi	Change values in Trino's hudi.properties file.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Change values in Trino's jmx.properties file.	Not available.
trino-connector-kafka	Change values in Trino's kafka.properties file.	Not available.
trino-connector-localfile	Change values in Trino's localfile.properties file.	Not available.
trino-connector-memory	Change values in Trino's memory.properties file.	Not available.
trino-connector-mongodb	Change values in Trino's mongodb.properties file.	Not available.
trino-connector-mysql	Change values in Trino's mysql.properties file.	Not available.
trino-connector-postgresql	Change values in Trino's postgresql.properties file.	Not available.
trino-connector-raptor	Change values in Trino's raptor.properties file.	Not available.
trino-connector-redis	Change values in Trino's redis.properties file.	Not available.
trino-connector-redshift	Change values in Trino's redshift.properties file.	Not available.
trino-connector-tpch	Change values in Trino's tpch.properties file.	Not available.
trino-connector-tpcds	Change values in Trino's tpcds.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-logback	Change values in kms-logba ck.xml file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j2	Change values in Spark's log4j2.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.

Classifications	Description	Reconfiguration Actions
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restart Oozie and HiveServe r2.
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts MapReduce -HistoryServer.
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Livy Server and MapReduce-HistoryS erver.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

6.15.0 change log

Change log for 6.15.0 release and release notes

Date	Event	Description
2023-11-17	Docs publication	Amazon EMR 6.15.0 release notes first published
2023-11-17	Deployment complete	Amazon EMR 6.15.0 fully deployed to all supported Regions
2023-11-13	Initial release	Amazon EMR 6.15.0 first deployed to initial commercia l Regions

Amazon EMR release 6.14.0

6.14.0 application versions

The following applications are supported in this release: Delta, Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, Iceberg, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Trino, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.14.0	emr-6.13.0	emr-6.12.0	emr-6.11.1
AWS SDK for Java	1.12.543	1.12.513	1.12.490	1.12.446
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.12.15	2.12.15	2.12.15	2.12.15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.4.0	2.4.0	2.4.0	2.2.0
Flink	1.17.1	1.17.0	1.17.0	1.16.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.17	2.4.17	2.4.17	2.4.15
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.13.1-amzn-2	0.13.1-amzn-1	0.13.1-amzn-0	0.13.0-amzn-0
Hue	4.11.0	4.11.0	4.11.0	4.11.0
Iceberg	1.3.1-amzn-0	1.3.0-amzn-1	1.3.0-amzn-0	1.2.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1

	emr-6.14.0	emr-6.13.0	emr-6.12.0	emr-6.11.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3	5.1.3	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.281	0.281	0.281	0.279
Spark	3.4.1	3.4.1	3.4.0	3.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.11.0
Tez	0.10.2	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	422	414	414	410
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

6.14.0 release notes

The following release notes include information for Amazon EMR release 6.14.0. Changes are relative to 6.13.0. For information on the release timeline, see the 6.14.0 change log.

New features

- Amazon EMR 6.14.0 supports Apache Spark 3.4.1, Apache Spark RAPIDS 23.06.0-amzn-2, Flink 1.17.1, Iceberg 1.3.1, and Trino 422.
- Amazon EMR managed scaling is now available in the `ap-southeast-3` Asia Pacific (Jakarta) Region for clusters that you create with Amazon EMR 6.14.0 and higher.

Changes, enhancements, and resolved issues

- The 6.14.0 release optimizes log management with Amazon EMR running on Amazon EC2. As a result, you might see a slight reduction in storage costs for your cluster logs.
- The 6.14.0 release improves the scaling workflow to account for different core instances that have a substantial variation in size for their Amazon EBS volumes. This improvement applies to core nodes only; scale-down operations for task nodes aren't affected.
- The 6.14.0 release improves the way that Amazon EMR interacts with open-source applications such as Apache Hadoop YARN ResourceManager and HDFS NameNode. This improvement reduces the risk of operational delays with cluster scaling, and mitigates startup failures that occur due to connectivity issues with the open-source applications.
- The 6.14.0 release optimizes application installation at cluster launch. This improves the cluster startup times for certain combinations of Amazon EMR applications.
- The 6.14.0 release fixes an issue where cluster scale-down operations might stall when a cluster that's running in a VPC with a custom domain encounters a core or task node restart.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see [Using the default Amazon Linux AMI for Amazon EMR](#).

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong),

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
			Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240223.0	4.14.336	March 8, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 131.0	4.14.336	February 14, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240124.0	4.14.336	February 7, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 109.0	4.14.334	January 24, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) ,Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia), Canada West (Calgary)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 218.0	4.14.330	January 2, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 206.0	4.14.330	December 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231116.0	4.14.328	December 11, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 101.0	4.14.327	November 17, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230906.0	4.14.322	September 11, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

6.14.0 default Java versions

Amazon EMR releases 6.12.0 and higher support all applications with Amazon Corretto 8 by default, except for Trino. For Trino, Amazon EMR supports Amazon Corretto 17 by default starting with Amazon EMR release 6.9.0. Amazon EMR also supports some applications with Amazon Corretto 11 and 17. Those applications are listed in the following table. If you want to change the default JVM on your cluster, follow the instructions in Configure applications to use a specific Java Virtual Machine for each application that runs on the cluster. You can only use one Java runtime version for a cluster. Amazon EMR doesn't support running different nodes or applications on different runtime versions on the same cluster.

While Amazon EMR supports both Amazon Corretto 11 and 17 on Apache Spark, Apache Hadoop, and Apache Hive, performance might regress for some workloads when you use these versions of Corretto. We recommend that you test your workloads before you change defaults.

The following table shows the default Java versions for applications in Amazon EMR 6.14.0:

Application	Java / Amazon Corretto version (default is bold)
Delta	17, 11, **8**
Flink	11, **8**
Ganglia	**8**
HBase	11, **8**
HCatalog	17, 11, **8**
Hadoop	17, 11, **8**
Hive	17, 11, **8**
Hudi	17, 11, **8**
Iceberg	17, 11, **8**
Livy	17, 11, **8**
Oozie	17, 11, **8**
Phoenix	**8**
PrestoDB	**8**

Application	Java / Amazon Corretto version (default is bold)
Spark	17, 11, **8**
Spark RAPIDS	17, 11, **8**
Sqoop	**8**
Tez	17, 11, **8**
Trino	**17**
Zeppelin	**8**
Pig	**8**
Zookeeper	**8**

6.14.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK

Component	Version	Description
delta	2.4.0	Delta lake is an open table format for huge analytic datasets
delta-standalone-connectors	0.6.0	Delta Connectors provide different runtimes to integrate Delta Lake with engines like Flink, Hive and Presto.
emr-ddb	5.1.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.7.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.11.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.7.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.28.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.7.0	EMR S3Select Connector
emr-wal-cli	1.1.0	Cli used for emrwal list/dele tion.
emrfs	2.59.0	Amazon S3 connector for Hadoop ecosystem applicati ons.

Component	Version	Description
flink-client	1.17.1-amzn-0	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.17.1-amzn-0	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	3.3.3-amzn-6	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.3.3-amzn-6	HDFS node-level service for storing blocks.
hadoop-hdfs-library	3.3.3-amzn-6	HDFS command-line client and library
hadoop-hdfs-namenode	3.3.3-amzn-6	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	3.3.3-amzn-6	HDFS service for managing the Hadoop filesystem journal on HA clusters.

Component	Version	Description
hadoop-httpfs-server	3.3.3-amzn-6	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.3.3-amzn-6	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.3.3-amzn-6	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.3.3-amzn-6	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.3.3-amzn-6	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.3.3-amzn-6	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	2.4.17-amzn-2	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.4.17-amzn-2	Service for serving one or more HBase regions.
hbase-client	2.4.17-amzn-2	HBase command-line client.
hbase-rest-server	2.4.17-amzn-2	Service providing a RESTful HTTP endpoint for HBase.

Component	Version	Description
hbase-thrift-server	2.4.17-amzn-2	Service providing a Thrift endpoint to HBase.
hbase-operator-tools	2.4.17-amzn-2	Repair tool for Apache HBase clusters.
hcatalog-client	3.1.3-amzn-7	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.3-amzn-7	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.3-amzn-7	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.3-amzn-7	Hive command line client.
hive-hbase	3.1.3-amzn-7	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-7	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	3.1.3-amzn-7	Service for accepting Hive queries as web requests.
hudi	0.13.1-amzn-2	Incremental processing framework to power data pipeline at low latency and high efficiency.

Component	Version	Description
hudi-presto	0.13.1-amzn-2	Bundle library for running Presto with Hudi.
hudi-trino	0.13.1-amzn-2	Bundle library for running Trino with Hudi.
hudi-spark	0.13.1-amzn-2	Bundle library for running Spark with Hudi.
hue-server	4.11.0	Web application for analyzing data using Hadoop ecosystem applications
iceberg	1.3.1-amzn-0	Apache Iceberg is an open table format for huge analytic datasets
jupyterhub	1.5.0	Multi-user server for Jupyter notebooks
livy-server	0.7.1-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.9.1	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68+	MariaDB database server.
nvidia-cuda	11.8.0	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.

Component	Version	Description
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.7.0	Open Source Computer Vision Library.
phoenix-library	5.1.3	The phoenix libraries for server and client
phoenix-connectors	5.1.3	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.3	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.281-amzn-2	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.281-amzn-2	Service for executing pieces of a query.
presto-client	0.281-amzn-2	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
trino-coordinator	422-amzn-0	Service for accepting queries and managing query execution among trino-wor kers.

Component	Version	Description
trino-worker	422-amzn-0	Service for executing pieces of a query.
trino-client	422-amzn-0	Trino command-line client which is installed on an HA cluster's stand-by masters where Trino server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.4.1-amzn-1	Spark command-line clients.
spark-history-server	3.4.1-amzn-1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	3.4.1-amzn-1	In-memory execution engine for YARN.
spark-yarn-slave	3.4.1-amzn-1	Apache Spark libraries needed by YARN slaves.
spark-rapids	23.06.0-amzn-2	Nvidia Spark RAPIDS plugin that accelerates Apache Spark with GPUs.
sqoop-client	1.4.7	Apache Sqoop command-line client.

Component	Version	Description
tensorflow	2.11.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.10.2-amzn-5	The tez YARN application and libraries.
tez-on-worker	0.10.2-amzn-5	The tez YARN application and libraries for worker nodes.
webserver	2.4.41+	Apache HTTP server.
zeppelin-server	0.10.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.5.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.5.10	ZooKeeper command line client.

6.14.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-6.14.0 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN

Classifications	Description	Reconfiguration Actions
		services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-ses sion.properties settings for Kubernetes/Yarn session.	Restarts Flink history server.
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Restarts Flink history server.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts PhoenixQu eryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classifications	Description	Reconfiguration Actions
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	This classification should not be reconfigured.
hdfs-env	Change values in the HDFS environment.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat server.

Classifications	Description	Reconfiguration Actions
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat server.
hive	Amazon EMR-curated settings for Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.
hive-env	Change values in the Hive environment.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Not available.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.

Classifications	Description	Reconfiguration Actions
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Change Hadoop's KMS java home	Not available.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Change values in the Hudi environment.	Not available.
hudi-defaults	Change values in Hudi's hudi-defaults.conf file.	Not available.

Classifications	Description	Reconfiguration Actions
iceberg-defaults	Change values in Iceberg's iceberg-defaults.conf file.	Not available.
delta-defaults	Change values in Delta's delta-defaults.conf file.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j2	Change Livy log4j2.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
oozie-env	Change values in Oozie's environment.	Restarts Oozie.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hb ase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServe r.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.pro perties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoDB)
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoDB)

Classifications	Description	Reconfiguration Actions
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoDB)
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-lakeformation	Change values in Presto's lakeformation.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
trino-log	Change values in Trino's log.properties file.	Restarts Trino-Server (for Trino)
trino-config	Change values in Trino's config.properties file.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Change values in Trino's password-authenticator.prop erties file.	Restarts Trino-Server (for Trino)
trino-env	Change values in Trino's trino-env.sh file.	Restarts Trino-Server (for Trino)
trino-node	Change values in Trino's node.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
trino-connector-blackhole	Change values in Trino's blackhole.properties file.	Not available.
trino-connector-cassandra	Change values in Trino's cassandra.properties file.	Not available.
trino-connector-delta	Change values in Trino's delta.properties file.	Restarts Trino-Server (for Trino)
trino-connector-hive	Change values in Trino's hive.properties file.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Change values in Trino's exchange-manager.properties file.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Change values in Trino's iceberg.properties file.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Change values in Trino's hudi.properties file.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Change values in Trino's jmx.properties file.	Not available.
trino-connector-kafka	Change values in Trino's kafka.properties file.	Not available.
trino-connector-localfile	Change values in Trino's localfile.properties file.	Not available.
trino-connector-memory	Change values in Trino's memory.properties file.	Not available.
trino-connector-mongodb	Change values in Trino's mongodb.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
trino-connector-mysql	Change values in Trino's mysql.properties file.	Not available.
trino-connector-postgresql	Change values in Trino's postgresql.properties file.	Not available.
trino-connector-raptor	Change values in Trino's raptor.properties file.	Not available.
trino-connector-redis	Change values in Trino's redis.properties file.	Not available.
trino-connector-redshift	Change values in Trino's redshift.properties file.	Not available.
trino-connector-tpch	Change values in Trino's tpch.properties file.	Not available.
trino-connector-tpcds	Change values in Trino's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-logback	Change values in kms-logba ck.xml file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.

Classifications	Description	Reconfiguration Actions
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j2	Change values in Spark's log4j2.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restart Oozie and HiveServer2.
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Livy Server and MapReduce-HistoryS erver.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

6.14.0 change log

Change log for 6.14.0 release and release notes

Date	Event	Description
2023-11-02	Deployment complete	Amazon EMR 6.14.0 fully deployed to all supported Regions
2023-10-10	Docs publication	Amazon EMR 6.14.0 release notes first published
2023-10-04	Initial release	Amazon EMR 6.14.0 first deployed to initial commercia l Regions

Amazon EMR release 6.13.0

6.13.0 application versions

The following applications are supported in this release: Delta, Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, Iceberg, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Trino, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.13.0	emr-6.12.0	emr-6.11.1	emr-6.11.0
AWS SDK for Java	1.12.513	1.12.490	1.12.446	1.12.446
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.12.15	2.12.15	2.12.15	2.12.15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.4.0	2.4.0	2.2.0	2.2.0
Flink	1.17.0	1.17.0	1.16.0	1.16.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.17	2.4.17	2.4.15	2.4.15

	emr-6.13.0	emr-6.12.0	emr-6.11.1	emr-6.11.0
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.13.1-amzn-1	0.13.1-amzn-0	0.13.0-amzn-0	0.13.0-amzn-0
Hue	4.11.0	4.11.0	4.11.0	4.11.0
Iceberg	1.3.0-amzn-1	1.3.0-amzn-0	1.2.0-amzn-0	1.2.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.281	0.281	0.279	0.279
Spark	3.4.1	3.4.0	3.3.2	3.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.11.0
Tez	0.10.2	0.10.2	0.10.2	0.10.2

	emr-6.13.0	emr-6.12.0	emr-6.11.1	emr-6.11.0
Trino (PrestoSQL)	414	414	410	410
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

6.13.0 release notes

The following release notes include information for Amazon EMR release 6.13.0. Changes are relative to 6.12.0. For information on the release timeline, see the 6.13.0 change log.

New features

- Amazon EMR 6.13.0 supports Apache Spark 3.4.1, Apache Spark RAPIDS 23.06.0-amzn-1, CUDA Toolkit 11.8.0, and JupyterHub 1.5.0.

Changes, enhancements, and resolved issues

- The 6.13.0 release improves the Amazon EMR log management daemon to ensure that all logs are uploaded at a regular cadence to Amazon S3 when a cluster termination command is issued. This facilitates faster cluster terminations.
- The 6.13.0 release enhances Amazon EMR log management capabilities to ensure consistent and timely upload of all log files to Amazon S3. This especially benefits long-running EMR clusters.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 816.0	4.14.350	August 21, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 809.0	4.14.349	August 20, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240719.0	4.14.348	July 25, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
			(Tel Aviv), Canada West (Calgary)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 223.0	4.14.336	March 8, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240131.0	4.14.336	February 14, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240124.0	4.14.336	February 7, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 109.0	4.14.334	January 24, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231218.0	4.14.330	January 2, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231206.0	4.14.330	December 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231116.0	4.14.328	December 11, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 101.0	4.14.327	November 16, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231020.1	4.14.326	November 7, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023012.1	4.14.326	October 26, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 926.0	4.14.322	October 19, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazoı Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 906.0	4.14.322	October 4, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 808.0	4.14.320	August 24, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

6.13.0 default Java versions

Amazon EMR releases 6.12.0 and higher support all applications with Amazon Corretto 8 by default, except for Trino. For Trino, Amazon EMR supports Amazon Corretto 17 by default starting with Amazon EMR release 6.9.0. Amazon EMR also supports some applications with Amazon Corretto 11 and 17. Those applications are listed in the following table. If you want to change the default JVM on your cluster, follow the instructions in Configure applications to use a specific Java Virtual Machine for each application that runs on the cluster. You can only use one Java runtime version for a cluster. Amazon EMR doesn't support running different nodes or applications on different runtime versions on the same cluster.

While Amazon EMR supports both Amazon Corretto 11 and 17 on Apache Spark, Apache Hadoop, and Apache Hive, performance might regress for some workloads when you use these versions of Corretto. We recommend that you test your workloads before you change defaults.

The following table shows the default Java versions for applications in Amazon EMR 6.13.0:

Application	Java / Amazon Corretto version (default is bold)
Delta	17, 11, **8**
Flink	11, **8**
Ganglia	**8**
HBase	11, **8**
HCatalog	17, 11, **8**
Hadoop	17, 11, **8**
Hive	17, 11, **8**
Hudi	17, 11, **8**
Iceberg	17, 11, **8**
Livy	17, 11, **8**
Oozie	17, 11, **8**
Phoenix	**8**
PrestoDB	**8**

Application	Java / Amazon Corretto version (default is bold)
Spark	17, 11, **8**
Spark RAPIDS	17, 11, **8**
Sqoop	**8**
Tez	17, 11, **8**
Trino	**17**
Zeppelin	**8**
Pig	**8**
Zookeeper	**8**

6.13.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK

Component	Version	Description
delta	2.4.0	Delta lake is an open table format for huge analytic datasets
delta-standalone-connectors	0.6.0	Delta Connectors provide different runtimes to integrate Delta Lake with engines like Flink, Hive and Presto.
emr-ddb	5.1.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.6.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.10.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.7.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.27.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.6.0	EMR S3Select Connector
emr-wal-cli	1.1.0	Cli used for emrwal list/dele tion.
emrfs	2.58.0	Amazon S3 connector for Hadoop ecosystem applicati ons.

Component	Version	Description
flink-client	1.17.0	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.17.0	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	3.3.3-amzn-5	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.3.3-amzn-5	HDFS node-level service for storing blocks.
hadoop-hdfs-library	3.3.3-amzn-5	HDFS command-line client and library
hadoop-hdfs-namenode	3.3.3-amzn-5	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	3.3.3-amzn-5	HDFS service for managing the Hadoop filesystem journal on HA clusters.

Component	Version	Description
hadoop-httpfs-server	3.3.3-amzn-5	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.3.3-amzn-5	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.3.3-amzn-5	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.3.3-amzn-5	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.3.3-amzn-5	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.3.3-amzn-5	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	2.4.17-amzn-1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.4.17-amzn-1	Service for serving one or more HBase regions.
hbase-client	2.4.17-amzn-1	HBase command-line client.
hbase-rest-server	2.4.17-amzn-1	Service providing a RESTful HTTP endpoint for HBase.

Component	Version	Description
hbase-thrift-server	2.4.17-amzn-1	Service providing a Thrift endpoint to HBase.
hbase-operator-tools	2.4.17-amzn-1	Repair tool for Apache HBase clusters.
hcatalog-client	3.1.3-amzn-6	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.3-amzn-6	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.3-amzn-6	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.3-amzn-6	Hive command line client.
hive-hbase	3.1.3-amzn-6	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-6	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	3.1.3-amzn-6	Service for accepting Hive queries as web requests.
hudi	0.13.1-amzn-1	Incremental processing framework to power data pipeline at low latency and high efficiency.

Component	Version	Description
hudi-presto	0.13.1-amzn-1	Bundle library for running Presto with Hudi.
hudi-trino	0.13.1-amzn-1	Bundle library for running Trino with Hudi.
hudi-spark	0.13.1-amzn-1	Bundle library for running Spark with Hudi.
hue-server	4.11.0	Web application for analyzing data using Hadoop ecosystem applications
iceberg	1.3.0-amzn-1	Apache Iceberg is an open table format for huge analytic datasets
jupyterhub	1.5.0	Multi-user server for Jupyter notebooks
livy-server	0.7.1-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.9.1	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68+	MariaDB database server.
nvidia-cuda	11.8.0	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.

Component	Version	Description
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.7.0	Open Source Computer Vision Library.
phoenix-library	5.1.3	The phoenix libraries for server and client
phoenix-connectors	5.1.3	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.3	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.281-amzn-1	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.281-amzn-1	Service for executing pieces of a query.
presto-client	0.281-amzn-1	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
trino-coordinator	414-amzn-1	Service for accepting queries and managing query execution among trino-wor kers.

Component	Version	Description
trino-worker	414-amzn-1	Service for executing pieces of a query.
trino-client	414-amzn-1	Trino command-line client which is installed on an HA cluster's stand-by masters where Trino server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.4.1-amzn-0	Spark command-line clients.
spark-history-server	3.4.1-amzn-0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	3.4.1-amzn-0	In-memory execution engine for YARN.
spark-yarn-slave	3.4.1-amzn-0	Apache Spark libraries needed by YARN slaves.
spark-rapids	23.06.0-amzn-1	Nvidia Spark RAPIDS plugin that accelerates Apache Spark with GPUs.
sqoop-client	1.4.7	Apache Sqoop command-line client.

Component	Version	Description
tensorflow	2.11.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.10.2-amzn-4	The tez YARN application and libraries.
tez-on-worker	0.10.2-amzn-4	The tez YARN application and libraries for worker nodes.
webserver	2.4.41+	Apache HTTP server.
zeppelin-server	0.10.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.5.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.5.10	ZooKeeper command line client.

6.13.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-6.13.0 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN

Classifications	Description	Reconfiguration Actions
		services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-session.properties settings for Kubernetes/Yarn session.	Restarts Flink history server.
flink-log4j-cli	Change Flink log4j-cli.properties settings.	Restarts Flink history server.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classifications	Description	Reconfiguration Actions
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	This classification should not be reconfigured.
hdfs-env	Change values in the HDFS environment.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat server.

Classifications	Description	Reconfiguration Actions
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat server.
hive	Amazon EMR-curated settings for Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.
hive-env	Change values in the Hive environment.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Not available.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.

Classifications	Description	Reconfiguration Actions
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Change Hadoop's KMS java home	Not available.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Change values in the Hudi environment.	Not available.
hudi-defaults	Change values in Hudi's hudi-defaults.conf file.	Not available.

Classifications	Description	Reconfiguration Actions
iceberg-defaults	Change values in Iceberg's iceberg-defaults.conf file.	Not available.
delta-defaults	Change values in Delta's delta-defaults.conf file.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j2	Change Livy log4j2.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
oozie-env	Change values in Oozie's environment.	Restarts Oozie.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hb ase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServe r.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.pro perties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoDB)
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoDB)

Classifications	Description	Reconfiguration Actions
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoDB)
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-lakeformation	Change values in Presto's lakeformation.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
trino-log	Change values in Trino's log.properties file.	Restarts Trino-Server (for Trino)
trino-config	Change values in Trino's config.properties file.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Change values in Trino's password-authenticator.prop erties file.	Restarts Trino-Server (for Trino)
trino-env	Change values in Trino's trino-env.sh file.	Restarts Trino-Server (for Trino)
trino-node	Change values in Trino's node.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
trino-connector-blackhole	Change values in Trino's blackhole.properties file.	Not available.
trino-connector-cassandra	Change values in Trino's cassandra.properties file.	Not available.
trino-connector-delta	Change values in Trino's delta.properties file.	Restarts Trino-Server (for Trino)
trino-connector-hive	Change values in Trino's hive.properties file.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Change values in Trino's exchange-manager.properties file.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Change values in Trino's iceberg.properties file.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Change values in Trino's hudi.properties file.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Change values in Trino's jmx.properties file.	Not available.
trino-connector-kafka	Change values in Trino's kafka.properties file.	Not available.
trino-connector-localfile	Change values in Trino's localfile.properties file.	Not available.
trino-connector-memory	Change values in Trino's memory.properties file.	Not available.
trino-connector-mongodb	Change values in Trino's mongodb.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
trino-connector-mysql	Change values in Trino's mysql.properties file.	Not available.
trino-connector-postgresql	Change values in Trino's postgresql.properties file.	Not available.
trino-connector-raptor	Change values in Trino's raptor.properties file.	Not available.
trino-connector-redis	Change values in Trino's redis.properties file.	Not available.
trino-connector-redshift	Change values in Trino's redshift.properties file.	Not available.
trino-connector-tpch	Change values in Trino's tpch.properties file.	Not available.
trino-connector-tpcds	Change values in Trino's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-logback	Change values in kms-logba ck.xml file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.

Classifications	Description	Reconfiguration Actions
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j2	Change values in Spark's log4j2.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restart Oozie and HiveServe r2.
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts MapReduce -HistoryServer.

Classifications	Description	Reconfiguration Actions
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

6.13.0 change log

Change log for 6.13.0 release and release notes

Date	Event	Description
2023-09-23	Deployment complete	Amazon EMR 6.13.0 fully deployed to all supported Regions
2023-09-12	Docs publication	Amazon EMR 6.13.0 release notes first published
2023-09-01	Initial release	Amazon EMR 6.13.0 first deployed to initial commercial Regions

Amazon EMR release 6.12.0

6.12.0 application versions

The following applications are supported in this release: Delta, Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, Iceberg, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Trino, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.12.0	emr-6.11.1	emr-6.11.0	emr-6.10.1
AWS SDK for Java	1.12.490	1.12.446	1.12.446	1.12.397
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.12.15	2.12.15	2.12.15	2.12.15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.4.0	2.2.0	2.2.0	2.2.0
Flink	1.17.0	1.16.0	1.16.0	1.16.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.17	2.4.15	2.4.15	2.4.15

	emr-6.12.0	emr-6.11.1	emr-6.11.0	emr-6.10.1
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.13.1-amzn-0	0.13.0-amzn-0	0.13.0-amzn-0	0.12.2-amzn-0
Hue	4.11.0	4.11.0	4.11.0	4.10.0
Iceberg	1.3.0-amzn-0	1.2.0-amzn-0	1.2.0-amzn-0	1.1.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.5.0
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.281	0.279	0.279	0.278
Spark	3.4.0	3.3.2	3.3.2	3.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.11.0
Tez	0.10.2	0.10.2	0.10.2	0.10.2

	emr-6.12.0	emr-6.11.1	emr-6.11.0	emr-6.10.1
Trino (PrestoSQL)	414	410	410	403
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

6.12.0 release notes

The following release notes include information for Amazon EMR release 6.12.0. Changes are relative to 6.11.0. For information on the release timeline, see the 6.12.0 change log.

New features

- Amazon EMR 6.12.0 supports Apache Spark 3.4.0, Apache Spark RAPIDS 23.06.0-amzn-0, CUDA 11.8.0, Apache Hudi 0.13.1-amzn-0, Apache Iceberg 1.3.0-amzn-0, Trino 414, and PrestoDB 0.281.
- Amazon EMR releases 6.12.0 and higher support LDAP integration with Apache Livy, Apache Hive through HiveServer2 (HS2), Trino, Presto, and Hue. You can also install Apache Spark and Apache Hadoop on an EMR cluster that uses 6.12.0 or higher and configure them to use LDAP. For more information, see Use Active Directory or LDAP servers for authentication with Amazon EMR.

Changes, enhancements, and resolved issues

- Amazon EMR releases 6.12.0 and higher provide Java 11 runtime support for Flink. For more information, see Configure Flink to run with Java 11.
- The 6.12.0 release adds a new retry mechanism to the cluster scaling workflow for EMR clusters that run Presto or Trino. This improvement reduces the risk that cluster resizing will indefinitely stall due to a single failed resize operation. It also improves cluster utilization, because your cluster scales up and down faster.
- The 6.12.0 release fixes an issue where cluster scale-down operations might stall when a core node that is undergoing graceful decommissioning turns unhealthy for any reason before it fully decommissions.

- The 6.12.0 release improves cluster scale-down logic so that your cluster doesn't attempt a scale-down of core nodes below the HDFS replication factor setting for the cluster. This aligns with your data redundancy requirements, and reduces the chance that a scaling operation might stall.
- The 6.12.0 release enhances the performance and efficiency of the health monitoring service for Amazon EMR by increasing the speed at which it logs state changes for instances. This improvement reduces the chance of degraded performance for cluster nodes that are running multiple custom client tools or third-party applications.
- The 6.12.0 release improves the performance of the on-cluster log management daemon for Amazon EMR. As a result, there is less chance for degraded performance with EMR clusters that run steps with high concurrency.
- With Amazon EMR release 6.12.0, the log management daemon has been upgraded to identify all logs that are in active use with open file handles on the local instance storage, and the associated processes. This upgrade ensures that Amazon EMR properly deletes the files and reclaims storage space after the logs are archived to Amazon S3.
- The 6.12.0 release includes a log-management daemon enhancement that deletes empty, unused steps directories in the local cluster file system. An excessively large number of empty directories can degrade the performance of Amazon EMR daemons and result in disk over-utilization.
- The 6.12.0 release enables log rotation for YARN Timeline Server logs. This minimizes disk over-utilization scenarios, especially for long-running clusters.
- The default root volume size has increased to 15 GB in Amazon EMR 6.10.0 and higher. Earlier releases have default root volume size of 10 GB.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see [Using the default Amazon Linux AMI for Amazon EMR](#).

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 816.0	4.14.350	August 21, 2024	US East (N. Virginia) , US East (Ohio), US

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
			West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024809.0	4.14.349	August 20, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240719.0	4.14.348	July 25, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
			(Tel Aviv), Canada West (Calgary)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 223.0	4.14.336	March 8, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 131.0	4.14.336	February 14, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia), Canada West (Calgary)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 124.0	4.14.336	February 7, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 109.0	4.14.334	January 24, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231218.0	4.14.330	January 2, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231206.0	4.14.330	December 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelease Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 116.0	4.14.328	December 11, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231101.0	4.14.327	November 16, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 020.1	4.14.326	November 7, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231012.1	4.14.326	October 26, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 926.0	4.14.322	October 19, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230906.0	4.14.322	October 4, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 822.0	4.14.322	Augest 30, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 808.0	4.14.320	August 24, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 727.0	4.14.320	August 14, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230719.0	4.14.320	August 2, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230628.0	4.14.318	July 12, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central)

6.12.0 default Java versions

Amazon EMR releases 6.12.0 and higher support all applications with Amazon Corretto 8 by default, except for Trino. For Trino, Amazon EMR supports Amazon Corretto 17 by default starting with Amazon EMR release 6.9.0. Amazon EMR also supports some applications with Amazon Corretto 11 and 17. Those applications are listed in the following table. If you want to change the default JVM on your cluster, follow the instructions in Configure applications to use a specific Java Virtual Machine for each application that runs on the cluster. You can only use one Java runtime version for a cluster. Amazon EMR doesn't support running different nodes or applications on different runtime versions on the same cluster.

While Amazon EMR supports both Amazon Corretto 11 and 17 on Apache Spark, Apache Hadoop, and Apache Hive, performance might regress for some workloads when you use these versions of Corretto. We recommend that you test your workloads before you change defaults.

The following table shows the default Java versions for applications in Amazon EMR 6.12.0:

Application	Java / Amazon Corretto version (default is bold)
Delta	17, 11, **8**
Flink	11, **8**
Ganglia	**8**
HBase	11, **8**
HCatalog	17, 11, **8**
Hadoop	17, 11, **8**
Hive	17, 11, **8**
Hudi	17, 11, **8**
Iceberg	17, 11, **8**
Livy	17, 11, **8**
Oozie	17, 11, **8**
Phoenix	**8**
PrestoDB	**8**
Spark	17, 11, **8**

Application	Java / Amazon Corretto version (default is bold)
Spark RAPIDS	17, 11, **8**
Sqoop	**8**
Tez	17, 11, **8**
Trino	**17**
Zeppelin	**8**
Pig	**8**
Zookeeper	**8**

6.12.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK

Component	Version	Description
delta	2.4.0	Delta lake is an open table format for huge analytic datasets
delta-standalone-connectors	0.6.0	Delta Connectors provide different runtimes to integrate Delta Lake with engines like Flink, Hive and Presto.
emr-ddb	5.1.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.5.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.9.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.7.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.26.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.5.0	EMR S3Select Connector
emr-wal-cli	1.1.0	Cli used for emrwal list/dele tion.
emrfs	2.57.0	Amazon S3 connector for Hadoop ecosystem applicati ons.

Component	Version	Description
flink-client	1.17.0	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.17.0	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	3.3.3-amzn-4	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.3.3-amzn-4	HDFS node-level service for storing blocks.
hadoop-hdfs-library	3.3.3-amzn-4	HDFS command-line client and library
hadoop-hdfs-namenode	3.3.3-amzn-4	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	3.3.3-amzn-4	HDFS service for managing the Hadoop filesystem journal on HA clusters.

Component	Version	Description
hadoop-httpfs-server	3.3.3-amzn-4	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.3.3-amzn-4	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.3.3-amzn-4	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.3.3-amzn-4	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.3.3-amzn-4	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.3.3-amzn-4	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	2.4.17-amzn-0	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.4.17-amzn-0	Service for serving one or more HBase regions.
hbase-client	2.4.17-amzn-0	HBase command-line client.
hbase-rest-server	2.4.17-amzn-0	Service providing a RESTful HTTP endpoint for HBase.

Component	Version	Description
hbase-thrift-server	2.4.17-amzn-0	Service providing a Thrift endpoint to HBase.
hbase-operator-tools	2.4.17-amzn-0	Repair tool for Apache HBase clusters.
hcatalog-client	3.1.3-amzn-5	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.3-amzn-5	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.3-amzn-5	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.3-amzn-5	Hive command line client.
hive-hbase	3.1.3-amzn-5	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-5	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	3.1.3-amzn-5	Service for accepting Hive queries as web requests.
hudi	0.13.1-amzn-0	Incremental processing framework to power data pipeline at low latency and high efficiency.

Component	Version	Description
hudi-presto	0.13.1-amzn-0	Bundle library for running Presto with Hudi.
hudi-trino	0.13.1-amzn-0	Bundle library for running Trino with Hudi.
hudi-spark	0.13.1-amzn-0	Bundle library for running Spark with Hudi.
hue-server	4.11.0	Web application for analyzing data using Hadoop ecosystem applications
iceberg	1.3.0-amzn-0	Apache Iceberg is an open table format for huge analytic datasets
jupyterhub	1.4.1	Multi-user server for Jupyter notebooks
livy-server	0.7.1-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.9.1	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68+	MariaDB database server.
nvidia-cuda	11.8.0	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.

Component	Version	Description
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.7.0	Open Source Computer Vision Library.
phoenix-library	5.1.3	The phoenix libraries for server and client
phoenix-connectors	5.1.3	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.3	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.281-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.281-amzn-0	Service for executing pieces of a query.
presto-client	0.281-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
trino-coordinator	414-amzn-0	Service for accepting queries and managing query execution among trino-wor kers.

Component	Version	Description
trino-worker	414-amzn-0	Service for executing pieces of a query.
trino-client	414-amzn-0	Trino command-line client which is installed on an HA cluster's stand-by masters where Trino server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.4.0-amzn-0	Spark command-line clients.
spark-history-server	3.4.0-amzn-0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	3.4.0-amzn-0	In-memory execution engine for YARN.
spark-yarn-slave	3.4.0-amzn-0	Apache Spark libraries needed by YARN slaves.
spark-rapids	23.06.0-amzn-0	Nvidia Spark RAPIDS plugin that accelerates Apache Spark with GPUs.
sqoop-client	1.4.7	Apache Sqoop command-line client.

Component	Version	Description
tensorflow	2.11.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.10.2-amzn-3	The tez YARN application and libraries.
tez-on-worker	0.10.2-amzn-3	The tez YARN application and libraries for worker nodes.
webserver	2.4.41+	Apache HTTP server.
zeppelin-server	0.10.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.5.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.5.10	ZooKeeper command line client.

6.12.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-6.12.0 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN

Classifications	Description	Reconfiguration Actions
		services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-ses sion.properties settings for Kubernetes/Yarn session.	Restarts Flink history server.
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Restarts Flink history server.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts PhoenixQu eryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classifications	Description	Reconfiguration Actions
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	This classification should not be reconfigured.
hdfs-env	Change values in the HDFS environment.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat server.

Classifications	Description	Reconfiguration Actions
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat server.
hive	Amazon EMR-curated settings for Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.
hive-env	Change values in the Hive environment.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Not available.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.

Classifications	Description	Reconfiguration Actions
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Change Hadoop's KMS java home	Not available.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Change values in the Hudi environment.	Not available.
hudi-defaults	Change values in Hudi's hudi-defaults.conf file.	Not available.

Classifications	Description	Reconfiguration Actions
iceberg-defaults	Change values in Iceberg's iceberg-defaults.conf file.	Not available.
delta-defaults	Change values in Delta's delta-defaults.conf file.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j2	Change Livy log4j2.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
oozie-env	Change values in Oozie's environment.	Restarts Oozie.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hb ase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServe r.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.pro perties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoDB)
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoDB)

Classifications	Description	Reconfiguration Actions
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoDB)
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-lakeformation	Change values in Presto's lakeformation.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
trino-log	Change values in Trino's log.properties file.	Restarts Trino-Server (for Trino)
trino-config	Change values in Trino's config.properties file.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Change values in Trino's password-authenticator.prop erties file.	Restarts Trino-Server (for Trino)
trino-env	Change values in Trino's trino-env.sh file.	Restarts Trino-Server (for Trino)
trino-node	Change values in Trino's node.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
trino-connector-blackhole	Change values in Trino's blackhole.properties file.	Not available.
trino-connector-cassandra	Change values in Trino's cassandra.properties file.	Not available.
trino-connector-delta	Change values in Trino's delta.properties file.	Restarts Trino-Server (for Trino)
trino-connector-hive	Change values in Trino's hive.properties file.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Change values in Trino's exchange-manager.properties file.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Change values in Trino's iceberg.properties file.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Change values in Trino's hudi.properties file.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Change values in Trino's jmx.properties file.	Not available.
trino-connector-kafka	Change values in Trino's kafka.properties file.	Not available.
trino-connector-localfile	Change values in Trino's localfile.properties file.	Not available.
trino-connector-memory	Change values in Trino's memory.properties file.	Not available.
trino-connector-mongodb	Change values in Trino's mongodb.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
trino-connector-mysql	Change values in Trino's mysql.properties file.	Not available.
trino-connector-postgresql	Change values in Trino's postgresql.properties file.	Not available.
trino-connector-raptor	Change values in Trino's raptor.properties file.	Not available.
trino-connector-redis	Change values in Trino's redis.properties file.	Not available.
trino-connector-redshift	Change values in Trino's redshift.properties file.	Not available.
trino-connector-tpch	Change values in Trino's tpch.properties file.	Not available.
trino-connector-tpcds	Change values in Trino's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-logback	Change values in kms-logba ck.xml file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.

Classifications	Description	Reconfiguration Actions
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j2	Change values in Spark's log4j2.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restart Oozie and HiveServe r2.
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts MapReduce -HistoryServer.

Classifications	Description	Reconfiguration Actions
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

6.12.0 change log

Change log for 6.12.0 release and release notes

Date	Event	Description
2023-07-27	Update documentation	Update the Java options for 6.12 and add Oozie tutorial to update JVM
2023-07-21	Deployment complete	Amazon EMR 6.12.0 fully deployed to all supported Regions
2023-07-21	Docs publication	Amazon EMR 6.12.0 release notes first published

Date	Event	Description
2023-07-12	Initial release	Amazon EMR 6.12.0 first deployed to initial commercia l Regions

Amazon EMR release 6.11.1

6.11.1 application versions

The following applications are supported in this release: Delta, Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, Iceberg, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Trino, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.11.1	emr-6.11.0	emr-6.10.1	emr-6.10.0
AWS SDK for Java	1.12.446	1.12.446	1.12.397	1.12.397
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.12.15	2.12.15	2.12.15	2.12.15
AmazonClo udWatchAgent	-	-	-	-

	emr-6.11.1	emr-6.11.0	emr-6.10.1	emr-6.10.0
Delta	2.2.0	2.2.0	2.2.0	2.2.0
Flink	1.16.0	1.16.0	1.16.0	1.16.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.15	2.4.15	2.4.15	2.4.15
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.13.0-amzn-0	0.13.0-amzn-0	0.12.2-amzn-0	0.12.2-amzn-0
Hue	4.11.0	4.11.0	4.10.0	4.10.0
Iceberg	1.2.0-amzn-0	1.2.0-amzn-0	1.1.0-amzn-0	1.1.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.4.1	1.4.1	1.5.0	1.5.0
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.279	0.279	0.278	0.278
Spark	3.3.2	3.3.2	3.3.1	3.3.1

	emr-6.11.1	emr-6.11.0	emr-6.10.1	emr-6.10.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.11.0
Tez	0.10.2	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	410	410	403	403
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

6.11.1 release notes

The following release notes include information for Amazon EMR release 6.11.1. Changes are relative to 6.11.0. For information on the release timeline, see the 6.11.1 change log.

Changes, enhancements, and resolved issues

- Due to lock contention, a node can enter into a deadlock if it's added or removed at the same time that it attempts to decommission. As a result, the Hadoop Resource Manager (YARN) becomes unresponsive, and affects all the incoming and currently-running containers.
- This release includes a change that allows high-availability clusters to recover from failed state after restart.
- This release includes security fixes for Hue and HBase.
- This release fixes an issue where clusters that are running workloads on Spark with Amazon EMR might silently receive incorrect results with `contains`, `startsWith`, `endsWith`, and `like`. This issue occurs when you use the expressions on partitioned fields that have metadata in the Amazon EMR Hive3 Metastore Server (HMS).
- This release fixes an issue with throttling on the Glue side when there are no user-defined functions (UDF).
- This release fixes an issue that deletes container logs by the node log aggregation service before log pusher can push them to S3 in case of YARN decommissioning.

- This release fixes an issue with FairShare Scheduler metrics when Node Label is enabled for Hadoop.
- This release fixes an issue that impacted Spark performance when you set a default `true` value for the `spark.yarn.heterogeneousExecutors.enabled` config in `spark-defaults.conf`.
- This release fixes an issue with Reduce Task failing to read shuffle data. The issue caused Hive query failures with a corrupted memory error.
- This release adds a new retry mechanism to the cluster scaling workflow for EMR clusters that run Presto or Trino. This improvement reduces the risk that cluster resizing will indefinitely stall due to a single failed resize operation. It also improves cluster utilization, because your cluster scales up and down faster.
- This release improves cluster scale-down logic so that your cluster doesn't attempt a scale-down of core nodes below the HDFS replication factor setting for the cluster. This aligns with your data redundancy requirements, and reduces the chance that a scaling operation might stall.
- The log management daemon has been upgraded to identify all logs that are in active use with open file handles on the local instance storage, and the associated processes. This upgrade ensures that Amazon EMR properly deletes the files and reclaims storage space after the logs are archived to Amazon S3.
- This release includes a log-management daemon enhancement that deletes empty, unused steps directories in the local cluster file system. An excessively large number of empty directories can degrade the performance of Amazon EMR daemons and result in disk over-utilization.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

OsRelease Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240816.0	4.14.350	August 21, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
			West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 809.0	4.14.349	August 20, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 719.0	4.14.348	July 25, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 709.1	4.14.348	July 23, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Eu rope (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing) , China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 223.0	4.14.336	March 8, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 131.0	4.14.336	February 14, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia), Canada West (Calgary)

OsRelease Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240124.0	4.14.336	February 7, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 109.0	4.14.334	January 24, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia), Canada West (Calgary)

OsRelea Label (Amazoı Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 218.0	4.14.330	January 2, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsRelea Label (Amazoı Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 206.0	4.14.330	December 22, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsRelea Label (Amazoı Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 116.0	4.14.328	December 11, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023101.0	4.14.327	November 16, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 020.1	4.14.326	November 7, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 012.1	4.14.326	October 26, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 926.0	4.14.322	October 19, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230906.0	4.14.322	October 4, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 822.0	4.14.322	Augest 30, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 808.0	4.14.320	August 24, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230727.0	4.14.320	August 14, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

6.11.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.2.0	Delta lake is an open table format for huge analytic datasets
delta-standalone-connectors	0.6.0	Delta Connectors provide different runtimes to integrate Delta Lake with engines like Flink, Hive and Presto.
emr-ddb	5.1.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.8.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.7.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.25.0	Distributed copy application optimized for Amazon S3.

Component	Version	Description
emr-s3-select	2.4.0	EMR S3Select Connector
emr-wal-cli	1.1.0	Cli used for emrwal list/dele tion.
emrfs	2.56.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.16.0	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.16.0	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	3.3.3-amzn-3.1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.3.3-amzn-3.1	HDFS node-level service for storing blocks.

Component	Version	Description
hadoop-hdfs-library	3.3.3-amzn-3.1	HDFS command-line client and library
hadoop-hdfs-namenode	3.3.3-amzn-3.1	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	3.3.3-amzn-3.1	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	3.3.3-amzn-3.1	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.3.3-amzn-3.1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.3.3-amzn-3.1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.3.3-amzn-3.1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.3.3-amzn-3.1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.3.3-amzn-3.1	Service for retrieving current and historical information for YARN applications.

Component	Version	Description
hbase-hmaster	2.4.15-amzn-1.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.4.15-amzn-1.1	Service for serving one or more HBase regions.
hbase-client	2.4.15-amzn-1.1	HBase command-line client.
hbase-rest-server	2.4.15-amzn-1.1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	2.4.15-amzn-1.1	Service providing a Thrift endpoint to HBase.
hbase-operator-tools	2.4.15-amzn-1.1	Repair tool for Apache HBase clusters.
hcatalog-client	3.1.3-amzn-4.1	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.3-amzn-4.1	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.3-amzn-4.1	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.3-amzn-4.1	Hive command line client.
hive-hbase	3.1.3-amzn-4.1	Hive-hbase client.

Component	Version	Description
hive-metastore-server	3.1.3-amzn-4.1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	3.1.3-amzn-4.1	Service for accepting Hive queries as web requests.
hudi	0.13.0-amzn-0	Incremental processing framework to power data pipeline at low latency and high efficiency.
hudi-presto	0.13.0-amzn-0	Bundle library for running Presto with Hudi.
hudi-trino	0.13.0-amzn-0	Bundle library for running Trino with Hudi.
hudi-spark	0.13.0-amzn-0	Bundle library for running Spark with Hudi.
hue-server	4.11.0	Web application for analyzing data using Hadoop ecosystem applications
iceberg	1.2.0-amzn-0	Apache Iceberg is an open table format for huge analytic datasets
jupyterhub	1.4.1	Multi-user server for Jupyter notebooks
livy-server	0.7.1-incubating	REST interface for interacting with Apache Spark

Component	Version	Description
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.9.1	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68+	MariaDB database server.
nvidia-cuda	11.8.0	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	The phoenix libraries for server and client
phoenix-connectors	5.1.2	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.279-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.

Component	Version	Description
presto-worker	0.279-amzn-0	Service for executing pieces of a query.
presto-client	0.279-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
trino-coordinator	410-amzn-0	Service for accepting queries and managing query execution among trino-wor kers.
trino-worker	410-amzn-0	Service for executing pieces of a query.
trino-client	410-amzn-0	Trino command-line client which is installed on an HA cluster's stand-by masters where Trino server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.3.2-amzn-0.1	Spark command-line clients.
spark-history-server	3.3.2-amzn-0.1	Web UI for viewing logged events for the lifetime of a completed Spark application.

Component	Version	Description
spark-on-yarn	3.3.2-amzn-0.1	In-memory execution engine for YARN.
spark-yarn-slave	3.3.2-amzn-0.1	Apache Spark libraries needed by YARN slaves.
spark-rapids	23.02.0-amzn-0	Nvidia Spark RAPIDS plugin that accelerates Apache Spark with GPUs.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.11.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.10.2-amzn-2.1	The tez YARN application and libraries.
tez-on-worker	0.10.2-amzn-2.1	The tez YARN application and libraries for worker nodes.
webserver	2.4.41+	Apache HTTP server.
zeppelin-server	0.10.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.5.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.

Component	Version	Description
zookeeper-client	3.5.10	ZooKeeper command line client.

6.11.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-6.11.1 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad

Classifications	Description	Reconfiguration Actions
		ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-ses sion.properties settings for Kubernetes/Yarn session.	Restarts Flink history server.

Classifications	Description	Reconfiguration Actions
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Restarts Flink history server.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts PhoenixQu eryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.

Classifications	Description	Reconfiguration Actions
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	This classification should not be reconfigured.
hdfs-env	Change values in the HDFS environment.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classifications	Description	Reconfiguration Actions
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat server.
hive	Amazon EMR-curated settings for Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
hive-env	Change values in the Hive environment.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Not available.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.

Classifications	Description	Reconfiguration Actions
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Change values in the Hudi environment.	Not available.
hudi-defaults	Change values in Hudi's hudi-defaults.conf file.	Not available.
iceberg-defaults	Change values in Iceberg's iceberg-defaults.conf file.	Not available.
delta-defaults	Change values in Delta's delta-defaults.conf file.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.

Classifications	Description	Reconfiguration Actions
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j2	Change Livy log4j2.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServer.

Classifications	Description	Reconfiguration Actions
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.pro perties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoDB)
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticat or	Change values in Presto's password-authenticator.prop erties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoDB)
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server (for PrestoDB)

Classifications	Description	Reconfiguration Actions
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-lakeformation	Change values in Presto's lakeformation.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
trino-log	Change values in Trino's log.properties file.	Restarts Trino-Server (for Trino)
trino-config	Change values in Trino's config.properties file.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Change values in Trino's password-authenticator.properties file.	Restarts Trino-Server (for Trino)
trino-env	Change values in Trino's trino-env.sh file.	Restarts Trino-Server (for Trino)
trino-node	Change values in Trino's node.properties file.	Not available.
trino-connector-blackhole	Change values in Trino's blackhole.properties file.	Not available.
trino-connector-cassandra	Change values in Trino's cassandra.properties file.	Not available.
trino-connector-delta	Change values in Trino's delta.properties file.	Restarts Trino-Server (for Trino)
trino-connector-hive	Change values in Trino's hive.properties file.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Change values in Trino's exchange-manager.properties file.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Change values in Trino's iceberg.properties file.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Change values in Trino's hudi.properties file.	Restarts Trino-Server (for Trino)

Classifications	Description	Reconfiguration Actions
trino-connector-jmx	Change values in Trino's jmx.properties file.	Not available.
trino-connector-kafka	Change values in Trino's kafka.properties file.	Not available.
trino-connector-localfile	Change values in Trino's localfile.properties file.	Not available.
trino-connector-memory	Change values in Trino's memory.properties file.	Not available.
trino-connector-mongodb	Change values in Trino's mongodb.properties file.	Not available.
trino-connector-mysql	Change values in Trino's mysql.properties file.	Not available.
trino-connector-postgresql	Change values in Trino's postgresql.properties file.	Not available.
trino-connector-raptor	Change values in Trino's raptor.properties file.	Not available.
trino-connector-redis	Change values in Trino's redis.properties file.	Not available.
trino-connector-redshift	Change values in Trino's redshift.properties file.	Not available.
trino-connector-tpch	Change values in Trino's tpch.properties file.	Not available.
trino-connector-tpcds	Change values in Trino's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.

Classifications	Description	Reconfiguration Actions
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-logback	Change values in kms-logba ck.xml file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j2	Change values in Spark's log4j2.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.

Classifications	Description	Reconfiguration Actions
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restart Oozie and HiveServe r2.
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts MapReduce -HistoryServer.
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Livy Server and MapReduce-HistoryS erver.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

6.11.1 change log

Change log for 6.11.1 release and release notes

Date	Event	Description
2023-08-30	Update release notes	Added several control-plane related fixes to the release notes
2023-08-21	Docs publication	Amazon EMR 6.11.1 release notes first published
2023-08-16	Deployment complete	Amazon EMR 6.11.1 fully deployed to all supported Regions
2023-08-04	Initial release	Amazon EMR 6.11.1 first deployed to limited commercial Regions

Amazon EMR release 6.11.0

6.11.0 application versions

The following applications are supported in this release: Delta, Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, Iceberg, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Trino, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.11.0	emr-6.10.1	emr-6.10.0	emr-6.9.1
AWS SDK for Java	1.12.446	1.12.397	1.12.397	1.12.170
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.12.15	2.12.15	2.12.15	2.12.15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.2.0	2.2.0	2.2.0	2.1.0
Flink	1.16.0	1.16.0	1.16.0	1.15.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.15	2.4.15	2.4.15	2.4.13
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.13.0-amzn-0	0.12.2-amzn-0	0.12.2-amzn-0	0.12.1-amzn-0
Hue	4.11.0	4.10.0	4.10.0	4.10.0
Iceberg	1.2.0-amzn-0	1.1.0-amzn-0	1.1.0-amzn-0	0.14.1-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.4.1	1.5.0	1.5.0	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1

	emr-6.11.0	emr-6.10.1	emr-6.10.0	emr-6.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.279	0.278	0.278	0.276
Spark	3.3.2	3.3.1	3.3.1	3.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.10.0
Tez	0.10.2	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	410	403	403	398
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

6.11.0 release notes

The following release notes include information for Amazon EMR release 6.11.0. Changes are relative to 6.10.0. For information on the release timeline, see the change log.

New features

- Amazon EMR 6.11.0 supports Apache Spark 3.3.2-amzn-0, Apache Spark RAPIDS 23.02.0-amzn-0, CUDA 11.8.0, Apache Hudi 0.13.0-amzn-0, Apache Iceberg 1.2.0-amzn-0, Trino 410-amzn-0, and PrestoDB 0.279-amzn-0.

Changes, enhancements, and resolved issues

- With Amazon EMR 6.11.0, the DynamoDB connector has been upgraded to version 5.0.0. Version 5.0.0 uses AWS SDK for Java 2.x. Previous releases used AWS SDK for Java 1.x. As a result of this upgrade, we strongly advise you to test your code before you use the DynamoDB connector with Amazon EMR 6.11.
- When the DynamoDB connector for Amazon EMR 6.11.0 calls the DynamoDB service, it uses the Region value that you provide for the `dynamodb.endpoint` property. We recommend that you also configure `dynamodb.region` when you use `dynamodb.endpoint`, and that both properties target the same AWS Region. If you use `dynamodb.endpoint` and you don't configure `dynamodb.region`, the DynamoDB connector for Amazon EMR 6.11.0 will return an invalid Region exception and attempt to reconcile your AWS Region information from the Amazon EC2 instance metadata service (IMDS). If the connector can't retrieve the Region from IMDS, it defaults to US East (N. Virginia) (`us-east-1`). The following error is an example of the invalid Region exception that you might get if you don't properly configure the `dynamodb.region` property: `error software.amazon.awssdk.services.dynamodb.model.DynamoDbException: Credential should be scoped to a valid region.` For more information on the classes that are affected by the AWS SDK for Java upgrade to 2.x, see the Upgrade AWS SDK for Java from 1.x to 2.x (#175) commit in the GitHub repo for the Amazon EMR - DynamoDB connector.
- This release fixes an issue where column data becomes `NULL` when you use Delta Lake to store Delta table data in Amazon S3 after column rename operation. For more information about this experimental feature in Delta Lake, see Column rename operation in the Delta Lake User Guide.
- The 6.11.0 release fixes an issue that might occur when you create an edge node by replicating one of the primary nodes from a cluster with multiple primary nodes. The replicated edge node could cause delays with scale-down operations, or result in high memory-utilization on the primary nodes. For more information on how to create an edge node to communicate with your EMR cluster, see Edge Node Creator in the `aws-samples` repo on GitHub.
- The 6.11.0 release improves the automation process that Amazon EMR uses to re-mount Amazon EBS volumes to an instance after a reboot.
- The 6.11.0 release fixes an issue that resulted in intermittent gaps in the Hadoop metrics that Amazon EMR publishes to Amazon CloudWatch.
- The 6.11.0 release fixes an issue with EMR clusters where an update to the YARN configuration file that contains the exclusion list of nodes for the cluster is interrupted due to disk over-

utilization. The incomplete update hinders future cluster scale-down operations. This release ensures that your cluster remains healthy, and that scaling operations work as expected.

- The default root volume size has increased to 15 GB in Amazon EMR 6.10.0 and higher. Earlier releases have default root volume size of 10 GB.
- Hadoop 3.3.3 introduced a change in YARN (YARN-9608) that keeps nodes where containers ran in a decommissioning state until the application completes. This change ensures that local data such as shuffle data doesn't get lost, and you don' need to re-run the job. This approach might also lead to underutilization of resources on clusters with or without managed scaling enabled.

  With Amazon EMR releases 6.11.0 and higher as well as 6.8.1, 6.9.1, and 6.10.1, the value of `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` is set to `false` in `yarn-site.xml` to resolve this issue.

  While the fix addresses the issues that were introduced by YARN-9608, it might cause Hive jobs to fail due to shuffle data loss on clusters that have managed scaling enabled. We've mitigated that risk in this release by also setting `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` for Hive workloads. This config is only available with Amazon EMR releases 6.11.0 and higher.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

  Note

  This release no longer gets automatic AMI updates since it has been succeeded by 1 more more patch releases. The patch release is denoted by the number after the second decimal point (`6.8.`*`1`*). To see if you're using the latest patch release, check the available releases in the *Release Guide*, or check the **Amazon EMR release** dropdown when you create a cluster in the console, or use the `ListReleaseLabels` API or `list-release-labels` CLI action. To get updates about new releases, subscribe to the RSS feed on the What's new? page.

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 816.0	4.14.350	August 21, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 809.0	4.14.349	August 20, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240719.0	4.14.348	July 25, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
			(Tel Aviv), Canada West (Calgary)
2.0.20230808.0	4.14.320	August 24, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), Israel (Tel Aviv)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 727.0	4.14.320	August 14, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230719.0	4.14.320	August 2, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230628.0	4.14.318	July 12, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Zurich), Europe (Milan), Europe (Spain), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 612.0	4.14.314	June 23, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Zurich), Europe (Milan), Europe (Spain), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 504.1	4.14.313	May 16, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central)

6.11.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the

most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.2.0	Delta lake is an open table format for huge analytic datasets
delta-standalone-connectors	0.6.0	Delta Connectors provide different runtimes to integrate Delta Lake with engines like Flink, Hive and Presto.
emr-ddb	5.1.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.8.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.7.0	Conda env for emr notebook which includes jupyter enterprise gateway

Component	Version	Description
emr-s3-dist-cp	2.25.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.4.0	EMR S3Select Connector
emr-wal-cli	1.1.0	Cli used for emrwal list/dele tion.
emrfs	2.56.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.16.0	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.16.0	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	3.3.3-amzn-3	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.

Component	Version	Description
hadoop-hdfs-datanode	3.3.3-amzn-3	HDFS node-level service for storing blocks.
hadoop-hdfs-library	3.3.3-amzn-3	HDFS command-line client and library
hadoop-hdfs-namenode	3.3.3-amzn-3	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	3.3.3-amzn-3	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	3.3.3-amzn-3	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.3.3-amzn-3	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.3.3-amzn-3	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.3.3-amzn-3	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.3.3-amzn-3	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.3.3-amzn-3	Service for retrieving current and historical information for YARN applications.

Component	Version	Description
hbase-hmaster	2.4.15-amzn-1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.4.15-amzn-1	Service for serving one or more HBase regions.
hbase-client	2.4.15-amzn-1	HBase command-line client.
hbase-rest-server	2.4.15-amzn-1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	2.4.15-amzn-1	Service providing a Thrift endpoint to HBase.
hbase-operator-tools	2.4.15-amzn-1	Repair tool for Apache HBase clusters.
hcatalog-client	3.1.3-amzn-4	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.3-amzn-4	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.3-amzn-4	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.3-amzn-4	Hive command line client.
hive-hbase	3.1.3-amzn-4	Hive-hbase client.

Component	Version	Description
hive-metastore-server	3.1.3-amzn-4	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	3.1.3-amzn-4	Service for accepting Hive queries as web requests.
hudi	0.13.0-amzn-0	Incremental processing framework to power data pipeline at low latency and high efficiency.
hudi-presto	0.13.0-amzn-0	Bundle library for running Presto with Hudi.
hudi-trino	0.13.0-amzn-0	Bundle library for running Trino with Hudi.
hudi-spark	0.13.0-amzn-0	Bundle library for running Spark with Hudi.
hue-server	4.11.0	Web application for analyzing data using Hadoop ecosystem applications
iceberg	1.2.0-amzn-0	Apache Iceberg is an open table format for huge analytic datasets
jupyterhub	1.4.1	Multi-user server for Jupyter notebooks
livy-server	0.7.1-incubating	REST interface for interacting with Apache Spark

Component	Version	Description
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.9.1	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68+	MariaDB database server.
nvidia-cuda	11.8.0	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	The phoenix libraries for server and client
phoenix-connectors	5.1.2	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.279-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.

Component	Version	Description
presto-worker	0.279-amzn-0	Service for executing pieces of a query.
presto-client	0.279-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
trino-coordinator	410-amzn-0	Service for accepting queries and managing query execution among trino-wor kers.
trino-worker	410-amzn-0	Service for executing pieces of a query.
trino-client	410-amzn-0	Trino command-line client which is installed on an HA cluster's stand-by masters where Trino server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.3.2-amzn-0	Spark command-line clients.
spark-history-server	3.3.2-amzn-0	Web UI for viewing logged events for the lifetime of a completed Spark application.

Component	Version	Description
spark-on-yarn	3.3.2-amzn-0	In-memory execution engine for YARN.
spark-yarn-slave	3.3.2-amzn-0	Apache Spark libraries needed by YARN slaves.
spark-rapids	23.02.0-amzn-0	Nvidia Spark RAPIDS plugin that accelerates Apache Spark with GPUs.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.11.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.10.2-amzn-2	The tez YARN application and libraries.
tez-on-worker	0.10.2-amzn-2	The tez YARN application and libraries for worker nodes.
webserver	2.4.41+	Apache HTTP server.
zeppelin-server	0.10.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.5.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.

Component	Version	Description
zookeeper-client	3.5.10	ZooKeeper command line client.

6.11.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-6.11.0 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad

Classifications	Description	Reconfiguration Actions
		ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-ses sion.properties settings for Kubernetes/Yarn session.	Restarts Flink history server.

Classifications	Description	Reconfiguration Actions
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Restarts Flink history server.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services SecondaryNamenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.

Classifications	Description	Reconfiguration Actions
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	This classification should not be reconfigured.
hdfs-env	Change values in the HDFS environment.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classifications	Description	Reconfiguration Actions
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat server.
hive	Amazon EMR-curated settings for Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
hive-env	Change values in the Hive environment.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Not available.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.

Classifications	Description	Reconfiguration Actions
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Change values in the Hudi environment.	Not available.
hudi-defaults	Change values in Hudi's hudi-defaults.conf file.	Not available.
iceberg-defaults	Change values in Iceberg's iceberg-defaults.conf file.	Not available.
delta-defaults	Change values in Delta's delta-defaults.conf file.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.

Classifications	Description	Reconfiguration Actions
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j2	Change Livy log4j2.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hb ase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServe r.

Classifications	Description	Reconfiguration Actions
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.pro perties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoDB)
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticat or	Change values in Presto's password-authenticator.prop erties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoDB)
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server (for PrestoDB)

Classifications	Description	Reconfiguration Actions
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-lakeformat ion	Change values in Presto's lakeformation.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
trino-log	Change values in Trino's log.properties file.	Restarts Trino-Server (for Trino)
trino-config	Change values in Trino's config.properties file.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Change values in Trino's password-authenticator.prop erties file.	Restarts Trino-Server (for Trino)
trino-env	Change values in Trino's trino-env.sh file.	Restarts Trino-Server (for Trino)
trino-node	Change values in Trino's node.properties file.	Not available.
trino-connector-blackhole	Change values in Trino's blackhole.properties file.	Not available.
trino-connector-cassandra	Change values in Trino's cassandra.properties file.	Not available.
trino-connector-delta	Change values in Trino's delta.properties file.	Restarts Trino-Server (for Trino)
trino-connector-hive	Change values in Trino's hive.properties file.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Change values in Trino's exchange-manager.properties file.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Change values in Trino's iceberg.properties file.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Change values in Trino's hudi.properties file.	Restarts Trino-Server (for Trino)

Classifications	Description	Reconfiguration Actions
trino-connector-jmx	Change values in Trino's jmx.properties file.	Not available.
trino-connector-kafka	Change values in Trino's kafka.properties file.	Not available.
trino-connector-localfile	Change values in Trino's localfile.properties file.	Not available.
trino-connector-memory	Change values in Trino's memory.properties file.	Not available.
trino-connector-mongodb	Change values in Trino's mongodb.properties file.	Not available.
trino-connector-mysql	Change values in Trino's mysql.properties file.	Not available.
trino-connector-postgresql	Change values in Trino's postgresql.properties file.	Not available.
trino-connector-raptor	Change values in Trino's raptor.properties file.	Not available.
trino-connector-redis	Change values in Trino's redis.properties file.	Not available.
trino-connector-redshift	Change values in Trino's redshift.properties file.	Not available.
trino-connector-tpch	Change values in Trino's tpch.properties file.	Not available.
trino-connector-tpcds	Change values in Trino's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.

Classifications	Description	Reconfiguration Actions
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-logback	Change values in kms-logba ck.xml file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j2	Change values in Spark's log4j2.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.

Classifications	Description	Reconfiguration Actions
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restart Oozie and HiveServe r2.
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts MapReduce -HistoryServer.
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Livy Server and MapReduce-HistoryS erver.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

6.11.0 change log

Change log for 6.11.0 release and release notes

Date	Event	Description
2023-08-21	Update	Fixed issue introduced with Hadoop 3.3.3.
2023-07-26	Update	New OS release labels `2.0.20230612.0` and `2.0.20230628.0` .
2023-06-09	Deployment complete	Amazon EMR 6.11.0 fully deployed to all supported Regions
2023-06-09	Docs publication	Amazon EMR 6.11.0 release notes first published
2023-06-08	Initial release	Amazon EMR 6.11.0 first deployed to initial commercia l Regions

Amazon EMR release 6.10.1

6.10.1 application versions

The following applications are supported in this release: Delta, Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, Iceberg, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Trino, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases

- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.10.1	emr-6.10.0	emr-6.9.1	emr-6.9.0
AWS SDK for Java	1.12.397	1.12.397	1.12.170	1.12.170
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.12.15	2.12.15	2.12.15	2.12.15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.2.0	2.2.0	2.1.0	2.1.0
Flink	1.16.0	1.16.0	1.15.2	1.15.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.15	2.4.15	2.4.13	2.4.13
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.12.2-amzn-0	0.12.2-amzn-0	0.12.1-amzn-0	0.12.1-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	1.1.0-amzn-0	1.1.0-amzn-0	0.14.1-amzn-0	0.14.1-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0

	emr-6.10.1	emr-6.10.0	emr-6.9.1	emr-6.9.0
JupyterHub	1.5.0	1.5.0	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.278	0.278	0.276	0.276
Spark	3.3.1	3.3.1	3.3.0	3.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.10.0	2.10.0
Tez	0.10.2	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	403	403	398	398
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

6.10.1 release notes

The following release notes include information for Amazon EMR release 6.10.1. Changes are relative to 6.10.0. For information on the release timeline, see the 6.10.1 change log.

Changes, enhancements, and resolved issues

- Due to lock contention, a node can enter into a deadlock if it's added or removed at the same time that it attempts to decommission. As a result, the Hadoop Resource Manager (YARN) becomes unresponsive, and affects all the incoming and currently-running containers.
- Hadoop 3.3.3 introduced a change in YARN (YARN-9608) that keeps nodes where containers ran in a decommissioning state until the application completes. This change ensures that local data such as shuffle data doesn't get lost, and you don' need to re-run the job. This approach might also lead to underutilization of resources on clusters with or without managed scaling enabled.

  With Amazon EMR releases 6.11.0 and higher as well as 6.8.1, 6.9.1, and 6.10.1, the value of `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` is set to `false` in `yarn-site.xml` to resolve this issue.

  While the fix addresses the issues that were introduced by YARN-9608, it might cause Hive jobs to fail due to shuffle data loss on clusters that have managed scaling enabled. We've mitigated that risk in this release by also setting `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` for Hive workloads. This config is only available with Amazon EMR releases 6.11.0 and higher.
- Metrics collector will not send any metrics to the control plane after failover of primary node in clusters with the instance groups configuration.
- This release includes a change that allows high-availability clusters to recover from failed state after restart.
- This release includes security fixes for Hue and HBase.
- This release fixes an issue where clusters that are running workloads on Spark with Amazon EMR might silently receive incorrect results with `contains`, `startsWith`, `endsWith`, and `like`. This issue occurs when you use the expressions on partitioned fields that have metadata in the Amazon EMR Hive3 Metastore Server (HMS).
- This release fixes an issue with throttling on the Glue side when there are no user-defined functions (UDF).
- This release fixes an issue that deletes container logs by the node log aggregation service before log pusher can push them to S3 in case of YARN decommissioning.
- This release fixes an issue with FairShare Scheduler metrics when Node Label is enabled for Hadoop.

- This release fixes an issue that impacted Spark performance when you set a default `true` value for the `spark.yarn.heterogeneousExecutors.enabled` config in `spark-defaults.conf`.
- This release fixes an issue with Reduce Task failing to read shuffle data. The issue caused Hive query failures with a corrupted memory error.
- This release adds a new retry mechanism to the cluster scaling workflow for EMR clusters that run Presto or Trino. This improvement reduces the risk that cluster resizing will indefinitely stall due to a single failed resize operation. It also improves cluster utilization, because your cluster scales up and down faster.
- This release improves cluster scale-down logic so that your cluster doesn't attempt a scale-down of core nodes below the HDFS replication factor setting for the cluster. This aligns with your data redundancy requirements, and reduces the chance that a scaling operation might stall.
- The log management daemon has been upgraded to identify all logs that are in active use with open file handles on the local instance storage, and the associated processes. This upgrade ensures that Amazon EMR properly deletes the files and reclaims storage space after the logs are archived to Amazon S3.
- This release includes a log-management daemon enhancement that deletes empty, unused steps directories in the local cluster file system. An excessively large number of empty directories can degrade the performance of Amazon EMR daemons and result in disk over-utilization.
- This release fixes an issue that might occur when you create an edge node by replicating one of the primary nodes from a cluster with multiple primary nodes. The replicated edge node could cause delays with scale-down operations, or result in high memory-utilization on the primary nodes. For more information on how to create an edge node to communicate with your EMR cluster, see Edge Node Creator in the `aws-samples` repo on GitHub.
- This release improves the automation process that Amazon EMR uses to re-mount Amazon EBS volumes to an instance after a reboot.
- This release fixes an issue that resulted in intermittent gaps in the Hadoop metrics that Amazon EMR publishes to Amazon CloudWatch.
- This release fixes an issue with EMR clusters where an update to the YARN configuration file that contains the exclusion list of nodes for the cluster is interrupted due to disk over-utilization. The incomplete update hinders future cluster scale-down operations. This release ensures that your cluster remains healthy, and that scaling operations work as expected.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2

release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240816.0	4.14.350	August 21, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240809.0	4.14.349	August 20, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 719.0	4.14.348	July 25, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsRelea Label (Amazoı Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 223.0	4.14.336	March 8, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia), Canada West (Calgary)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 131.0	4.14.336	February 14, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240124.0	4.14.336	February 7, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240109.0	4.14.334	January 24, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 218.0	4.14.330	January 2, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20231206.0	4.14.330	December 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 116.0	4.14.328	December 11, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20231101.0	4.14.327	November 16, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 020.1	4.14.326	November 7, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20231012.1	4.14.326	October 26, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 926.0	4.14.322	October 19, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230906.0	4.14.322	October 4, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 822.0	4.14.322	Augest 30, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230808.0	4.14.320	August 24, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230727.0	4.14.320	August 14, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

6.10.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.2.0	Delta lake is an open table format for huge analytic datasets
emr-ddb	4.16.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.3.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.7.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.7.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.24.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.3.0	EMR S3Select Connector
emr-wal-cli	1.0.0	Cli used for emrwal list/dele tion.

Component	Version	Description
emrfs	2.55.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.16.0	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.16.0	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	3.3.3-amzn-2.1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.3.3-amzn-2.1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	3.3.3-amzn-2.1	HDFS command-line client and library
hadoop-hdfs-namenode	3.3.3-amzn-2.1	HDFS service for tracking file names and block locations.

Component	Version	Description
hadoop-hdfs-journalnode	3.3.3-amzn-2.1	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	3.3.3-amzn-2.1	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.3.3-amzn-2.1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.3.3-amzn-2.1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.3.3-amzn-2.1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.3.3-amzn-2.1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.3.3-amzn-2.1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	2.4.15-amzn-0.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.4.15-amzn-0.1	Service for serving one or more HBase regions.
hbase-client	2.4.15-amzn-0.1	HBase command-line client.

Component	Version	Description
hbase-rest-server	2.4.15-amzn-0.1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	2.4.15-amzn-0.1	Service providing a Thrift endpoint to HBase.
hbase-operator-tools	2.4.15-amzn-0.1	Repair tool for Apache HBase clusters.
hcatalog-client	3.1.3-amzn-3.1	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.3-amzn-3.1	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.3-amzn-3.1	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.3-amzn-3.1	Hive command line client.
hive-hbase	3.1.3-amzn-3.1	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-3.1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	3.1.3-amzn-3.1	Service for accepting Hive queries as web requests.

Component	Version	Description
hudi	0.12.2-amzn-0	Incremental processing framework to power data pipeline at low latency and high efficiency.
hudi-presto	0.12.2-amzn-0	Bundle library for running Presto with Hudi.
hudi-trino	0.12.2-amzn-0	Bundle library for running Trino with Hudi.
hudi-spark	0.12.2-amzn-0	Bundle library for running Spark with Hudi.
hue-server	4.10.0	Web application for analyzing data using Hadoop ecosystem applications
iceberg	1.1.0-amzn-0	Apache Iceberg is an open table format for huge analytic datasets
jupyterhub	1.5.0	Multi-user server for Jupyter notebooks
livy-server	0.7.1-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.9.1	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68+	MariaDB database server.

Component	Version	Description
nvidia-cuda	11.8.0	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	The phoenix libraries for server and client
phoenix-connectors	6.0.0-SNAPSHOT	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	6.0.0	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.278.1-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.278.1-amzn-0	Service for executing pieces of a query.
presto-client	0.278.1-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.

Component	Version	Description
trino-coordinator	403-amzn-0	Service for accepting queries and managing query execution among trino-wor kers.
trino-worker	403-amzn-0	Service for executing pieces of a query.
trino-client	403-amzn-0	Trino command-line client which is installed on an HA cluster's stand-by masters where Trino server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.3.1-amzn-0.1	Spark command-line clients.
spark-history-server	3.3.1-amzn-0.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	3.3.1-amzn-0.1	In-memory execution engine for YARN.
spark-yarn-slave	3.3.1-amzn-0.1	Apache Spark libraries needed by YARN slaves.
spark-rapids	22.12.0-amzn-0	Nvidia Spark RAPIDS plugin that accelerates Apache Spark with GPUs.

Component	Version	Description
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.11.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.10.2-amzn-1.1	The tez YARN application and libraries.
tez-on-worker	0.10.2-amzn-1.1	The tez YARN application and libraries for worker nodes.
webserver	2.4.41+	Apache HTTP server.
zeppelin-server	0.10.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.5.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.5.10	ZooKeeper command line client.

6.10.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-6.10.1 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode,

Classifications	Description	Reconfiguration Actions
		SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-ses sion.properties settings for Kubernetes/Yarn session.	Restarts Flink history server.
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Restarts Flink history server.

Classifications	Description	Reconfiguration Actions
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services SecondaryNamenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.

Classifications	Description	Reconfiguration Actions
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	This classification should not be reconfigured.
hdfs-env	Change values in the HDFS environment.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classifications	Description	Reconfiguration Actions
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat server.
hive	Amazon EMR-curated settings for Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
hive-env	Change values in the Hive environment.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Not available.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.

Classifications	Description	Reconfiguration Actions
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Change values in the Hudi environment.	Not available.
hudi-defaults	Change values in Hudi's hudi-defaults.conf file.	Not available.
iceberg-defaults	Change values in Iceberg's iceberg-defaults.conf file.	Not available.
delta-defaults	Change values in Delta's delta-defaults.conf file.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.

Classifications	Description	Reconfiguration Actions
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j2	Change Livy log4j2.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hb ase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServe r.

Classifications	Description	Reconfiguration Actions
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoDB)
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoDB)
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server (for PrestoDB)

Classifications	Description	Reconfiguration Actions
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-lakeformat ion	Change values in Presto's lakeformation.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
trino-log	Change values in Trino's log.properties file.	Restarts Trino-Server (for Trino)
trino-config	Change values in Trino's config.properties file.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Change values in Trino's password-authenticator.properties file.	Restarts Trino-Server (for Trino)
trino-env	Change values in Trino's trino-env.sh file.	Restarts Trino-Server (for Trino)
trino-node	Change values in Trino's node.properties file.	Not available.
trino-connector-blackhole	Change values in Trino's blackhole.properties file.	Not available.
trino-connector-cassandra	Change values in Trino's cassandra.properties file.	Not available.
trino-connector-delta	Change values in Trino's delta.properties file.	Restarts Trino-Server (for Trino)
trino-connector-hive	Change values in Trino's hive.properties file.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Change values in Trino's exchange-manager.properties file.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Change values in Trino's iceberg.properties file.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Change values in Trino's hudi.properties file.	Restarts Trino-Server (for Trino)

Classifications	Description	Reconfiguration Actions
trino-connector-jmx	Change values in Trino's jmx.properties file.	Not available.
trino-connector-kafka	Change values in Trino's kafka.properties file.	Not available.
trino-connector-localfile	Change values in Trino's localfile.properties file.	Not available.
trino-connector-memory	Change values in Trino's memory.properties file.	Not available.
trino-connector-mongodb	Change values in Trino's mongodb.properties file.	Not available.
trino-connector-mysql	Change values in Trino's mysql.properties file.	Not available.
trino-connector-postgresql	Change values in Trino's postgresql.properties file.	Not available.
trino-connector-raptor	Change values in Trino's raptor.properties file.	Not available.
trino-connector-redis	Change values in Trino's redis.properties file.	Not available.
trino-connector-redshift	Change values in Trino's redshift.properties file.	Not available.
trino-connector-tpch	Change values in Trino's tpch.properties file.	Not available.
trino-connector-tpcds	Change values in Trino's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.

Classifications	Description	Reconfiguration Actions
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-logback	Change values in kms-logba ck.xml file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j2	Change values in Spark's log4j2.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.

Classifications	Description	Reconfiguration Actions
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restart Oozie and HiveServe r2.
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts MapReduce -HistoryServer.
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Livy Server and MapReduce-HistoryS erver.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

6.10.1 change log

Change log for 6.10.1 release and release notes

Date	Event	Description
2023-08-30	Update release notes	Added several control-plane related fixes to the release notes
2023-08-21	Docs publication	Amazon EMR 6.10.1 release notes first published
2023-08-16	Deployment complete	Amazon EMR 6.10.1 fully deployed to all supported Regions
2023-08-04	Initial release	Amazon EMR 6.10.1 first deployed to limited commercial Regions

Amazon EMR release 6.10.0

6.10.0 application versions

The following applications are supported in this release: Delta, Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, Iceberg, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Trino, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.10.0	emr-6.9.1	emr-6.9.0	emr-6.8.1
AWS SDK for Java	1.12.397	1.12.170	1.12.170	1.12.170
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.12.15	2.12.15	2.12.15	2.12.15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.2.0	2.1.0	2.1.0	-
Flink	1.16.0	1.15.2	1.15.2	1.15.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.15	2.4.13	2.4.13	2.4.12
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.2.1
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.12.2-amzn-0	0.12.1-amzn-0	0.12.1-amzn-0	0.11.1-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	1.1.0-amzn-0	0.14.1-amzn-0	0.14.1-amzn-0	0.14.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.1.0
JupyterHub	1.5.0	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1

	emr-6.10.0	emr-6.9.1	emr-6.9.0	emr-6.8.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.278	0.276	0.276	0.273
Spark	3.3.1	3.3.0	3.3.0	3.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.10.0	2.10.0	2.9.1
Tez	0.10.2	0.10.2	0.10.2	0.9.2
Trino (PrestoSQL)	403	398	398	388
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

6.10.0 release notes

The following release notes include information for Amazon EMR release 6.10.0. Changes are relative to 6.9.0. For information on the release timeline, see the change log.

New features

- Amazon EMR 6.10.0 supports Apache Spark 3.3.1, Apache Spark RAPIDS 22.12.0, CUDA 11.8.0, Apache Hudi 0.12.2-amzn-0, Apache Iceberg 1.1.0-amzn-0, Trino 403, and PrestoDB 0.278.1.
- Amazon EMR 6.10.0 includes a native Trino-Hudi connector that provides read access to data in Hudi tables. You can activate the connector with `trino-cli --catalog hudi`, and configure the connector for your requirements with `trino-connector-hudi`. The native integration with Amazon EMR means that you no longer need to use `trino-connector-hive` to query Hudi

tables. For a list of supported configurations with the new connector, see the Hudi connector page of the Trino documentation.

- Amazon EMR releases 6.10.0 and higher support Apache Zeppelin integration with Apache Flink. See Working with Flink jobs from Zeppelin in Amazon EMR for more information.

Known Issues

- Hadoop 3.3.3 introduced a change in YARN (YARN-9608) that keeps nodes where containers ran in a decommissioning state until the application completes. This change ensures that local data such as shuffle data doesn't get lost, and you don' need to re-run the job. This approach might also lead to underutilization of resources on clusters with or without managed scaling enabled.

  To work around this issue in Amazon EMR 6.10.0, you can set the value of `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` to `false` in `yarn-site.xml`. In Amazon EMR releases 6.11.0 and higher as well as 6.8.1, 6.9.1, and 6.10.1, the config is set to `false` by default to resolve this issue.

Changes, enhancements, and resolved issues

- Amazon EMR 6.10.0 removes the dependency on `minimal-json.jar` for the Amazon Redshift integration for Apache Spark, and automatically adds the required Spark-Redshift related jars to the executor class path for Spark: `spark-redshift.jar`, `spark-avro.jar`, and `RedshiftJDBC.jar`.
- The 6.10.0 release improves the on-cluster log management daemon to monitor additional log folders in your EMR cluster. This improvement minimizes disk over-utilization scenarios.
- The 6.10.0 release automatically restarts the on-cluster log management daemon when it stops. This improvement reduces the risk for nodes to appear unhealthy due to disk over-utilization.
- Amazon EMR 6.10.0 supports regional endpoints for EMRFS user mapping.
- The default root volume size has increased to 15 GB in Amazon EMR 6.10.0 and higher. Earlier releases have default root volume size of 10 GB.
- The 6.10.0 release fixes an issue that caused Spark jobs to stall when all remaining Spark executors are on a decommissioning host with the YARN resource manager.
- With Amazon EMR 6.6.0 through 6.9.x, INSERT queries with dynamic partition and an ORDER BY or SORT BY clause will always have two reducers. This issue is caused by OSS change HIVE-20703, which puts dynamic sort partition optimization under cost-based decision. If

your workload doesn't require sorting of dynamic partitions, we recommend that you set the `hive.optimize.sort.dynamic.partition.threshold` property to `-1` to disable the new feature and get the correctly calculated number of reducers. This issue is fixed in OSS Hive as part of HIVE-22269 and is fixed in Amazon EMR 6.10.0.

- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

> **Note**
>
> This release no longer gets automatic AMI updates since it has been succeeded by 1 more more patch releases. The patch release is denoted by the number after the second decimal point (`6.8.1`). To see if you're using the latest patch release, check the available releases in the *Release Guide*, or check the **Amazon EMR release** dropdown when you create a cluster in the console, or use the `ListReleaseLabels` API or `list-release-labels` CLI action. To get updates about new releases, subscribe to the RSS feed on the What's new? page.

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 816.0	4.14.350	August 21, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
			(Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 809.0	4.14.349	August 20, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 719.0	4.14.348	July 25, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230808.0	4.14.320	August 24, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230727.0	4.14.320	August 14, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230719.0	4.14.320	August 2, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230628.0	4.14.318	July 12, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Zurich), Europe (Milan), Europe (Spain), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 612.0	4.14.314	June 23, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Zurich), Europe (Milan), Europe (Spain), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230504.1	4.14.313	May 16, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Zurich), Europe (Milan), Europe (Spain), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230418.0	4.14.311	May 3, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Zurich), Europe (Milan), Europe (Spain), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230404.1	4.14.311	April 18, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE)
2.0.20230404.0	4.14.311	April 10, 2023	US East (N. Virginia), Europe (Paris)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 320.0	4.14.309	March 30, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230207.0	4.14.304	February 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE)

6.10.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.2.0	Delta lake is an open table format for huge analytic datasets
emr-ddb	4.16.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.3.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.7.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.7.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.24.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.3.0	EMR S3Select Connector
emr-wal-cli	1.0.0	Cli used for emrwal list/dele tion.

Component	Version	Description
emrfs	2.55.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.16.0	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.16.0	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	3.3.3-amzn-2	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.3.3-amzn-2	HDFS node-level service for storing blocks.
hadoop-hdfs-library	3.3.3-amzn-2	HDFS command-line client and library
hadoop-hdfs-namenode	3.3.3-amzn-2	HDFS service for tracking file names and block locations.

Component	Version	Description
hadoop-hdfs-journalnode	3.3.3-amzn-2	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	3.3.3-amzn-2	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.3.3-amzn-2	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.3.3-amzn-2	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.3.3-amzn-2	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.3.3-amzn-2	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.3.3-amzn-2	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	2.4.15-amzn-0	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.4.15-amzn-0	Service for serving one or more HBase regions.
hbase-client	2.4.15-amzn-0	HBase command-line client.

Component	Version	Description
hbase-rest-server	2.4.15-amzn-0	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	2.4.15-amzn-0	Service providing a Thrift endpoint to HBase.
hbase-operator-tools	2.4.15-amzn-0	Repair tool for Apache HBase clusters.
hcatalog-client	3.1.3-amzn-3	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.3-amzn-3	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.3-amzn-3	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.3-amzn-3	Hive command line client.
hive-hbase	3.1.3-amzn-3	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-3	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	3.1.3-amzn-3	Service for accepting Hive queries as web requests.

Component	Version	Description
hudi	0.12.2-amzn-0	Incremental processing framework to power data pipeline at low latency and high efficiency.
hudi-presto	0.12.2-amzn-0	Bundle library for running Presto with Hudi.
hudi-trino	0.12.2-amzn-0	Bundle library for running Trino with Hudi.
hudi-spark	0.12.2-amzn-0	Bundle library for running Spark with Hudi.
hue-server	4.10.0	Web application for analyzing data using Hadoop ecosystem applications
iceberg	1.1.0-amzn-0	Apache Iceberg is an open table format for huge analytic datasets
jupyterhub	1.5.0	Multi-user server for Jupyter notebooks
livy-server	0.7.1-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.9.1	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68+	MariaDB database server.

Component	Version	Description
nvidia-cuda	11.8.0	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	The phoenix libraries for server and client
phoenix-connectors	6.0.0-SNAPSHOT	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	6.0.0	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.278.1-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.278.1-amzn-0	Service for executing pieces of a query.
presto-client	0.278.1-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.

Component	Version	Description
trino-coordinator	403-amzn-0	Service for accepting queries and managing query execution among trino-wor kers.
trino-worker	403-amzn-0	Service for executing pieces of a query.
trino-client	403-amzn-0	Trino command-line client which is installed on an HA cluster's stand-by masters where Trino server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.3.1-amzn-0	Spark command-line clients.
spark-history-server	3.3.1-amzn-0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	3.3.1-amzn-0	In-memory execution engine for YARN.
spark-yarn-slave	3.3.1-amzn-0	Apache Spark libraries needed by YARN slaves.
spark-rapids	22.12.0-amzn-0	Nvidia Spark RAPIDS plugin that accelerates Apache Spark with GPUs.

Component	Version	Description
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.11.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.10.2-amzn-1	The tez YARN application and libraries.
tez-on-worker	0.10.2-amzn-1	The tez YARN application and libraries for worker nodes.
webserver	2.4.41+	Apache HTTP server.
zeppelin-server	0.10.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.5.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.5.10	ZooKeeper command line client.

6.10.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-6.10.0 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode,

Classifications	Description	Reconfiguration Actions
		SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-session.properties settings for Kubernetes/Yarn session.	Restarts Flink history server.
flink-log4j-cli	Change Flink log4j-cli.properties settings.	Restarts Flink history server.

Classifications	Description	Reconfiguration Actions
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services SecondaryNamenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.

Classifications	Description	Reconfiguration Actions
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	This classification should not be reconfigured.
hdfs-env	Change values in the HDFS environment.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classifications	Description	Reconfiguration Actions
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat server.
hive	Amazon EMR-curated settings for Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
hive-env	Change values in the Hive environment.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Not available.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.

Classifications	Description	Reconfiguration Actions
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Change values in the Hudi environment.	Not available.
hudi-defaults	Change values in Hudi's hudi-defaults.conf file.	Not available.
iceberg-defaults	Change values in Iceberg's iceberg-defaults.conf file.	Not available.
delta-defaults	Change values in Delta's delta-defaults.conf file.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.

Classifications	Description	Reconfiguration Actions
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j2	Change Livy log4j2.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServer.

Classifications	Description	Reconfiguration Actions
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.pro perties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoDB)
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticat or	Change values in Presto's password-authenticator.prop erties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoDB)
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server (for PrestoDB)

Classifications	Description	Reconfiguration Actions
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-lakeformat ion	Change values in Presto's lakeformation.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
trino-log	Change values in Trino's log.properties file.	Restarts Trino-Server (for Trino)
trino-config	Change values in Trino's config.properties file.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Change values in Trino's password-authenticator.properties file.	Restarts Trino-Server (for Trino)
trino-env	Change values in Trino's trino-env.sh file.	Restarts Trino-Server (for Trino)
trino-node	Change values in Trino's node.properties file.	Not available.
trino-connector-blackhole	Change values in Trino's blackhole.properties file.	Not available.
trino-connector-cassandra	Change values in Trino's cassandra.properties file.	Not available.
trino-connector-delta	Change values in Trino's delta.properties file.	Restarts Trino-Server (for Trino)
trino-connector-hive	Change values in Trino's hive.properties file.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Change values in Trino's exchange-manager.properties file.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Change values in Trino's iceberg.properties file.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Change values in Trino's hudi.properties file.	Restarts Trino-Server (for Trino)

Classifications	Description	Reconfiguration Actions
trino-connector-jmx	Change values in Trino's jmx.properties file.	Not available.
trino-connector-kafka	Change values in Trino's kafka.properties file.	Not available.
trino-connector-localfile	Change values in Trino's localfile.properties file.	Not available.
trino-connector-memory	Change values in Trino's memory.properties file.	Not available.
trino-connector-mongodb	Change values in Trino's mongodb.properties file.	Not available.
trino-connector-mysql	Change values in Trino's mysql.properties file.	Not available.
trino-connector-postgresql	Change values in Trino's postgresql.properties file.	Not available.
trino-connector-raptor	Change values in Trino's raptor.properties file.	Not available.
trino-connector-redis	Change values in Trino's redis.properties file.	Not available.
trino-connector-redshift	Change values in Trino's redshift.properties file.	Not available.
trino-connector-tpch	Change values in Trino's tpch.properties file.	Not available.
trino-connector-tpcds	Change values in Trino's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.

Classifications	Description	Reconfiguration Actions
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-logback	Change values in kms-logba ck.xml file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j2	Change values in Spark's log4j2.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.

Classifications	Description	Reconfiguration Actions
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restart Oozie and HiveServer2.
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

6.10.0 change log

Change log for 6.10.0 release and release notes

Date	Event	Description
2023-08-21	Update	Added a known issue introduced with Hadoop 3.3.3.
2023-07-26	Update	New OS release labels `2.0.20230612.0` and `2.0.20230628.0` .
2023-03-02	Deployment complete	Amazon EMR 6.10 fully deployed to all supported Regions
2023-03-02	Docs publication	Amazon EMR 6.10 release notes first published
2023-02-27	Initial release	Amazon EMR 6.10 deployed to limited commercial Regions

Amazon EMR release 6.9.1

6.9.1 application versions

The following applications are supported in this release: Delta, Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, Iceberg, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Trino, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases

- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.9.1	emr-6.9.0	emr-6.8.1	emr-6.8.0
AWS SDK for Java	1.12.170	1.12.170	1.12.170	1.12.170
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.12.15	2.12.15	2.12.15	2.12.15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.1.0	2.1.0	-	-
Flink	1.15.2	1.15.2	1.15.1	1.15.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.13	2.4.13	2.4.12	2.4.12
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.2.1	3.2.1
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.12.1-amzn-0	0.12.1-amzn-0	0.11.1-amzn-0	0.11.1-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	0.14.1-amzn-0	0.14.1-amzn-0	0.14.0-amzn-0	0.14.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.1.0	2.1.0

	emr-6.9.1	emr-6.9.0	emr-6.8.1	emr-6.8.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.276	0.276	0.273	0.273
Spark	3.3.0	3.3.0	3.3.0	3.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.10.0	2.10.0	2.9.1	2.9.1
Tez	0.10.2	0.10.2	0.9.2	0.9.2
Trino (PrestoSQL)	398	398	388	388
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

6.9.1 release notes

The following release notes include information for Amazon EMR release 6.9.1. Changes are relative to 6.9.0. For information on the release timeline, see the 6.9.1 change log.

Changes, enhancements, and resolved issues

- Hadoop 3.3.3 introduced a change in YARN (YARN-9608) that keeps nodes where containers ran in a decommissioning state until the application completes. This change ensures that local data such as shuffle data doesn't get lost, and you don' need to re-run the job. This approach might also lead to underutilization of resources on clusters with or without managed scaling enabled.

  With Amazon EMR releases 6.11.0 and higher as well as 6.8.1, 6.9.1, and 6.10.1, the value of `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` is set to `false` in `yarn-site.xml` to resolve this issue.

  While the fix addresses the issues that were introduced by YARN-9608, it might cause Hive jobs to fail due to shuffle data loss on clusters that have managed scaling enabled. We've mitigated that risk in this release by also setting `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` for Hive workloads. This config is only available with Amazon EMR releases 6.11.0 and higher.
- Metrics collector won't send any metrics to the control plane after failover of primary node in clusters with the instance groups configuration.
- This release eliminates retries on failed HTTP requests to metrics collector endpoints.
- This release includes a change that allows high-availability clusters to recover from failed state after restart.
- This release fixes an issue where large user-created UIDs caused overflow exceptions.
- This release fixes timeout issues with the Amazon EMR reconfiguration process.
- This release includes security fixes.
- This release fixes an issue where clusters that are running workloads on Spark with Amazon EMR might silently receive incorrect results with `contains`, `startsWith`, `endsWith`, and `like`. This issue occurs when you use the expressions on partitioned fields that have metadata in the Amazon EMR Hive3 Metastore Server (HMS).
- With Amazon EMR 6.6.0 through 6.9.x, INSERT queries with dynamic partition and an ORDER BY or SORT BY clause will always have two reducers. This issue is caused by OSS change HIVE-20703, which puts dynamic sort partition optimization under cost-based decision. If your workload doesn't require sorting of dynamic partitions, we recommend that you set the `hive.optimize.sort.dynamic.partition.threshold` property to `-1` to disable the new feature and get the correctly calculated number of reducers. This issue is fixed in OSS Hive as part of HIVE-22269 and is fixed in Amazon EMR 6.10.0.

- Hive might experience data loss when you use HDFS as a scratch directory and you have enabled merge small files, and the table contains static partition paths.
- This release fixes a performance issue with Hive if merge small files (disabled by default) is enabled at the end of ETL job.
- This release fixes an issue with throttling on the Glue side when there are no user-defined functions (UDF).
- This release fixes an issue that deletes container logs by the node log aggregation service before log pusher can push them to S3 in case of YARN decommissioning.
- This release fixes handling of compacted/archived files with persistent storefile tracking for HBase.
- This release fixes an issue that impacted Spark performance when you set a default `true` value for the `spark.yarn.heterogeneousExecutors.enabled` config in `spark-defaults.conf`.
- This release fixes an issue with Reduce Task failing to read shuffle data. The issue caused Hive query failures with a corrupted memory error.
- This release fixes an issue that caused the node provisioner to fail if the HDFS NameNode (NN) service was stuck in safemode during node replacement.
- This release adds a new retry mechanism to the cluster scaling workflow for EMR clusters that run Presto or Trino. This improvement reduces the risk that cluster resizing will indefinitely stall due to a single failed resize operation. It also improves cluster utilization, because your cluster scales up and down faster.
- This release improves cluster scale-down logic so that your cluster doesn't attempt a scale-down of core nodes below the HDFS replication factor setting for the cluster. This aligns with your data redundancy requirements, and reduces the chance that a scaling operation might stall.
- The log management daemon has been upgraded to identify all logs that are in active use with open file handles on the local instance storage, and the associated processes. This upgrade ensures that Amazon EMR properly deletes the files and reclaims storage space after the logs are archived to Amazon S3.
- This release includes a log-management daemon enhancement that deletes empty, unused steps directories in the local cluster file system. An excessively large number of empty directories can degrade the performance of Amazon EMR daemons and result in disk over-utilization.
- This release fixes an issue that might occur when you create an edge node by replicating one of the primary nodes from a cluster with multiple primary nodes. The replicated edge node could cause delays with scale-down operations, or result in high memory-utilization on the primary

nodes. For more information on how to create an edge node to communicate with your EMR cluster, see [Edge Node Creator] in the `aws-samples` repo on GitHub.

- This release improves the automation process that Amazon EMR uses to re-mount Amazon EBS volumes to an instance after a reboot.
- This release fixes an issue that resulted in intermittent gaps in the Hadoop metrics that Amazon EMR publishes to Amazon CloudWatch.
- This release fixes an issue with EMR clusters where an update to the YARN configuration file that contains the exclusion list of nodes for the cluster is interrupted due to disk over-utilization. The incomplete update hinders future cluster scale-down operations. This release ensures that your cluster remains healthy, and that scaling operations work as expected.
- This release improves the on-cluster log management daemon to monitor additional log folders in your EMR cluster. This improvement minimizes disk over-utilization scenarios.
- This release automatically restarts the on-cluster log management daemon when it stops. This improvement reduces the risk for nodes to appear unhealthy due to disk over-utilization.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see [Using the default Amazon Linux AMI for Amazon EMR].

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240816.0	4.14.350	August 21, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
			(Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240809.0	4.14.349	August 20, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240719.0	4.14.348	July 25, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 223.0	4.14.336	March 8, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West) , AWS GovCloud (US-East) , China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240131.0	4.14.336	February 14, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 124.0	4.14.336	February 7, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West) , AWS GovCloud (US-East) , China (Beijing), China (Ningxia), Canada West (Calgary)

OsRelease Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 109.0	4.14.334	January 24, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West) , AWS GovCloud (US-East) , China (Beijing), China (Ningxia), Canada West (Calgary)

OsRelea Label (Amazoı Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 218.0	4.14.330	January 2, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West) , AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20231206.0	4.14.330	December 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20231116.0	4.14.328	December 11, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20231101.0	4.14.327	November 16, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 020.1	4.14.326	November 7, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West) , AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20231012.1	4.14.326	October 26, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 926.0	4.14.322	October 19, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 906.0	4.14.322	October 4, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 822.0	4.14.322	Augest 30, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230808.0	4.14.320	August 24, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 727.0	4.14.320	August 14, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

6.9.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.1.0	Delta lake is an open table format for huge analytic datasets
emr-ddb	4.16.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.3.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.6.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.7.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.23.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.2.0	EMR S3Select Connector
emrfs	2.54.0	Amazon S3 connector for Hadoop ecosystem applicati ons.

Component	Version	Description
flink-client	1.15.2	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.15.2	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	3.3.3-amzn-1.1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.3.3-amzn-1.1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	3.3.3-amzn-1.1	HDFS command-line client and library
hadoop-hdfs-namenode	3.3.3-amzn-1.1	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	3.3.3-amzn-1.1	HDFS service for managing the Hadoop filesystem journal on HA clusters.

Component	Version	Description
hadoop-httpfs-server	3.3.3-amzn-1.1	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.3.3-amzn-1.1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.3.3-amzn-1.1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.3.3-amzn-1.1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.3.3-amzn-1.1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.3.3-amzn-1.1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	2.4.13-amzn-0.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.4.13-amzn-0.1	Service for serving one or more HBase regions.
hbase-client	2.4.13-amzn-0.1	HBase command-line client.
hbase-rest-server	2.4.13-amzn-0.1	Service providing a RESTful HTTP endpoint for HBase.

Component	Version	Description
hbase-thrift-server	2.4.13-amzn-0.1	Service providing a Thrift endpoint to HBase.
hbase-operator-tools	2.4.13-amzn-0.1	Repair tool for Apache HBase clusters.
hcatalog-client	3.1.3-amzn-2.1	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.3-amzn-2.1	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.3-amzn-2.1	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.3-amzn-2.1	Hive command line client.
hive-hbase	3.1.3-amzn-2.1	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-2.1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	3.1.3-amzn-2.1	Service for accepting Hive queries as web requests.
hudi	0.12.1-amzn-0	Incremental processing framework to power data pipeline at low latency and high efficiency.

Component	Version	Description
hudi-presto	0.12.1-amzn-0	Bundle library for running Presto with Hudi.
hudi-trino	0.12.1-amzn-0	Bundle library for running Trino with Hudi.
hudi-spark	0.12.1-amzn-0	Bundle library for running Spark with Hudi.
hue-server	4.10.0	Web application for analyzing data using Hadoop ecosystem applications
iceberg	0.14.1-amzn-0	Apache Iceberg is an open table format for huge analytic datasets
jupyterhub	1.4.1	Multi-user server for Jupyter notebooks
livy-server	0.7.1-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.9.1	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68+	MariaDB database server.
nvidia-cuda	11.7.0	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.

Component	Version	Description
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	The phoenix libraries for server and client
phoenix-connectors	6.0.0-SNAPSHOT	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	6.0.0	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.276-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.276-amzn-0	Service for executing pieces of a query.
presto-client	0.276-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
trino-coordinator	398-amzn-0	Service for accepting queries and managing query execution among trino-wor kers.

Component	Version	Description
trino-worker	398-amzn-0	Service for executing pieces of a query.
trino-client	398-amzn-0	Trino command-line client which is installed on an HA cluster's stand-by masters where Trino server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.3.0-amzn-1.1	Spark command-line clients.
spark-history-server	3.3.0-amzn-1.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	3.3.0-amzn-1.1	In-memory execution engine for YARN.
spark-yarn-slave	3.3.0-amzn-1.1	Apache Spark libraries needed by YARN slaves.
spark-rapids	22.08.0-amzn-0	Nvidia Spark RAPIDS plugin that accelerates Apache Spark with GPUs.
sqoop-client	1.4.7	Apache Sqoop command-line client.

Component	Version	Description
tensorflow	2.10.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.10.2-amzn-0.1	The tez YARN application and libraries.
webserver	2.4.41+	Apache HTTP server.
zeppelin-server	0.10.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.5.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.5.10	ZooKeeper command line client.

6.9.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-6.9.1 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN

Classifications	Description	Reconfiguration Actions
		services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-ses sion.properties settings for Kubernetes/Yarn session.	Restarts Flink history server.
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Restarts Flink history server.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts PhoenixQu eryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classifications	Description	Reconfiguration Actions
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	This classification should not be reconfigured.
hdfs-env	Change values in the HDFS environment.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat server.

Classifications	Description	Reconfiguration Actions
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat server.
hive	Amazon EMR-curated settings for Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.
hive-env	Change values in the Hive environment.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Not available.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.

Classifications	Description	Reconfiguration Actions
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Change values in the Hudi environment.	Not available.
hudi-defaults	Change values in Hudi's hudi-defaults.conf file.	Not available.
iceberg-defaults	Change values in Iceberg's iceberg-defaults.conf file.	Not available.

Classifications	Description	Reconfiguration Actions
delta-defaults	Change values in Delta's delta-defaults.conf file.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j2	Change Livy log4j2.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.

Classifications	Description	Reconfiguration Actions
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hb ase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServe r.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.pro perties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoDB)
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoDB)

Classifications	Description	Reconfiguration Actions
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoDB)
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-lakeformation	Change values in Presto's lakeformation.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
trino-log	Change values in Trino's log.properties file.	Restarts Trino-Server (for Trino)
trino-config	Change values in Trino's config.properties file.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Change values in Trino's password-authenticator.prop erties file.	Restarts Trino-Server (for Trino)
trino-env	Change values in Trino's trino-env.sh file.	Restarts Trino-Server (for Trino)
trino-node	Change values in Trino's node.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
trino-connector-blackhole	Change values in Trino's blackhole.properties file.	Not available.
trino-connector-cassandra	Change values in Trino's cassandra.properties file.	Not available.
trino-connector-delta	Change values in Trino's delta.properties file.	Restarts Trino-Server (for Trino)
trino-connector-hive	Change values in Trino's hive.properties file.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Change values in Trino's exchange-manager.properties file.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Change values in Trino's iceberg.properties file.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Change values in Trino's jmx.properties file.	Not available.
trino-connector-kafka	Change values in Trino's kafka.properties file.	Not available.
trino-connector-localfile	Change values in Trino's localfile.properties file.	Not available.
trino-connector-memory	Change values in Trino's memory.properties file.	Not available.
trino-connector-mongodb	Change values in Trino's mongodb.properties file.	Not available.
trino-connector-mysql	Change values in Trino's mysql.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
trino-connector-postgresql	Change values in Trino's postgresql.properties file.	Not available.
trino-connector-raptor	Change values in Trino's raptor.properties file.	Not available.
trino-connector-redis	Change values in Trino's redis.properties file.	Not available.
trino-connector-redshift	Change values in Trino's redshift.properties file.	Not available.
trino-connector-tpch	Change values in Trino's tpch.properties file.	Not available.
trino-connector-tpcds	Change values in Trino's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-log4j	Change values in kms-log4j .properties file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.

Classifications	Description	Reconfiguration Actions
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j2	Change values in Spark's log4j2.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restart Oozie and HiveServe r2.
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts MapReduce -HistoryServer.

Classifications	Description	Reconfiguration Actions
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

6.9.1 change log

Change log for 6.9.1 release and release notes

Date	Event	Description
2023-08-30	Update release notes	Added several control-plane related fixes to the release notes
2023-08-21	Docs publication	Amazon EMR 6.9.1 release notes first published
2023-08-16	Deployment complete	Amazon EMR 6.9.1 fully deployed to all supported Regions

Date	Event	Description
2023-08-04	Initial release	Amazon EMR 6.9.1 first deployed to limited commercial Regions

Amazon EMR release 6.9.0

6.9.0 application versions

The following applications are supported in this release: Delta, Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, Iceberg, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Trino, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.9.0	emr-6.8.1	emr-6.8.0	emr-6.7.0
AWS SDK for Java	1.12.170	1.12.170	1.12.170	1.12.170
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.12.15	2.12.15	2.12.15	2.12.15
AmazonCloudWatchAgent	-	-	-	-

	emr-6.9.0	emr-6.8.1	emr-6.8.0	emr-6.7.0
Delta	2.1.0	-	-	-
Flink	1.15.2	1.15.1	1.15.1	1.14.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.13	2.4.12	2.4.12	2.4.4
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.2.1	3.2.1	3.2.1
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.12.1-amzn-0	0.11.1-amzn-0	0.11.1-amzn-0	0.11.0-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	0.14.1-amzn-0	0.14.0-amzn-0	0.14.0-amzn-0	0.13.1-amzn-0
JupyterEnterpriseGateway	2.6.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.8.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.276	0.273	0.273	0.272
Spark	3.3.0	3.3.0	3.3.0	3.2.1

	emr-6.9.0	emr-6.8.1	emr-6.8.0	emr-6.7.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.10.0	2.9.1	2.9.1	2.4.1
Tez	0.10.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	398	388	388	378
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.0
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.7

6.9.0 release notes

The following release notes include information for Amazon EMR release 6.9.0. Changes are relative to Amazon EMR release 6.8.0. For information on the release timeline, see the change log.

New Features

- Amazon EMR release 6.9.0 supports Apache Spark RAPIDS 22.08.0, Apache Hudi 0.12.1, Apache Iceberg 0.14.1, Trino 398, and Tez 0.10.2.
- Amazon EMR release 6.9.0 includes a new open-source application, Delta Lake 2.1.0.
- The Amazon Redshift integration for Apache Spark is included in Amazon EMR releases 6.9.0 and later. Previously an open-source tool, the native integration is a Spark connector that you can use to build Apache Spark applications that read from and write to data in Amazon Redshift and Amazon Redshift Serverless. For more information, see Using Amazon Redshift integration for Apache Spark with Amazon EMR .
- Amazon EMR release 6.9.0 adds support for archiving logs to Amazon S3 during cluster scale-down. Previously, you could only archive log files to Amazon S3 during cluster termination. The new capability ensures that log files generated on the cluster persist on Amazon S3 even after the node is terminated. For more information, see Configure cluster logging and debugging.
- To support long running queries, Trino now includes a fault-tolerant execution mechanism. Fault-tolerant execution mitigates query failures by retrying failed queries or their component tasks. For more information, see Fault-tolerant execution in Trino.

- You can use Apache Flink on Amazon EMR for unified BATCH and STREAM processing of Apache Hive Tables or metadata of any Flink tablesource such as Iceberg, Kinesis or Kafka. You can specify the AWS Glue Data Catalog as the metastore for Flink using the AWS Management Console, AWS CLI, or Amazon EMR API. For more information, see Configuring Flink in Amazon EMR.
- You can now specify AWS Identity and Access Management (IAM) runtime roles and AWS Lake Formation-based access control for Apache Spark, Apache Hive, and Presto queries on Amazon EMR on EC2 clusters with Amazon SageMaker Studio. For more information, see Configure runtime roles for Amazon EMR steps.

Known Issues

- For Amazon EMR release 6.9.0, Trino does not work on clusters enabled for Apache Ranger. If you need to use Trino with Ranger, contact AWS Support.
- If you use the the Amazon Redshift integration for Apache Spark and have a time, timetz, timestamp, or timestamptz with microsecond precision in Parquet format, the connector rounds the time values to the nearest millisecond value. As a workaround, use the text unload format `unload_s3_format` parameter.
- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&‘()*+,-`. For more information, see UTF-8 encoding table and Unicode characters.

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

- Connections to Amazon EMR clusters from Amazon SageMaker Studio may intermittently fail with a **403 Forbidden** response code. This error happens when setup of the IAM role on the cluster takes longer than 60 seconds. As a workaround, you can install an Amazon EMR patch to enable retries and increase the timeout to a minimum of 300 seconds. Use the following steps to apply the bootstrap action when you launch your cluster.

  1. Download the bootstrap script and RPM files from the following Amazon S3 URIs.

     ```
     s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/gcsc/replace-rpms.sh
     s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/gcsc/emr-secret-agent-1.18.0-SNAPSHOT20221121212949.noarch.rpm
     ```

  2. Upload the files from the previous step to an Amazon S3 bucket that you own. The bucket must be in the same AWS Region where you plan to launch the cluster.
  3. Include the following bootstrap action when you launch your EMR cluster. Replace *bootstrap_URI* and *RPM_URI* with the corresponding URIs from Amazon S3.

     ```
     --bootstrap-actions "Path=bootstrap_URI,Args=[RPM_URI]"
     ```

- With Amazon EMR releases 5.36.0 and 6.6.0 through 6.9.0, `SecretAgent` and `RecordServer` service components may experience log data loss due to an incorrect file name pattern configuration in Log4j2 properties. The incorrect configuration causes the components to generate only one log file per day. When the rotation strategy occurs, it overwrites the existing file instead of generating a new log file as expected. As a workaround, use a bootstrap action to generate log files each hour and append an auto-increment integer in the file name to handle the rotation.

  For Amazon EMR 6.6.0 through 6.9.0 releases, use the following bootstrap action when you launch a cluster.

  ```
  --bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
  ```

  For Amazon EMR 5.36.0, use the following bootstrap action when you launch a cluster.

  ```
  --bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
  ```

- Apache Flink provides Native S3 FileSystem and Hadoop FileSystem Connectors, which let applications create a FileSink and write the data into Amazon S3. This FileSink fails with one of the following two exceptions.

```
java.lang.UnsupportedOperationException: Recoverable writers on Hadoop are only
 supported for HDFS
```

```
Caused by: java.lang.NoSuchMethodError:
 org.apache.hadoop.io.retry.RetryPolicies.retryOtherThanRemoteAndSaslException(Lorg/
apache/hadoop/io/retry/RetryPolicy;Ljava/util/Map;)Lorg/apache/hadoop/io/retry/
RetryPolicy;
                                        at
 org.apache.hadoop.yarn.client.RMProxy.createRetryPolicy(RMProxy.java:302) ~[hadoop-
yarn-common-3.3.3-amzn-0.jar:?]
```

As a workaround, you can install an Amazon EMR patch, which fixes the above issue in Flink. To apply the bootstrap action when you launch your cluster, complete the following steps.

1. Download the flink-rpm to your Amazon S3 bucket. Your RPM path is `s3://DOC-EXAMPLE-BUCKET/rpms/flink/`.
2. Download the bootstrap script and RPM files from Amazon S3 using the following URI. Replace *regionName* with the AWS Region where you plan to launch the cluster.

   ```
   s3://emr-data-access-control-regionName/customer-bootstrap-actions/gcsc/replace-
   rpms.sh
   ```

3. Hadoop 3.3.3 introduced a change in YARN (YARN-9608) that keeps nodes where containers ran in a decommissioning state until the application completes. This change ensures that local data such as shuffle data doesn't get lost, and you don' need to re-run the job. In Amazon EMR 6.8.0 and 6.9.0, this approach might also lead to underutilization of resources on clusters with or without managed scaling enabled.

   With Amazon EMR 6.10.0, there's a workaround for this issue to set the value of `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` to `false` in `yarn-site.xml`. In Amazon EMR releases 6.11.0 and higher as well as 6.8.1, 6.9.1, and 6.10.1, the config is set to `false` by default to resolve this issue.

Changes, Enhancements, and Resolved Issues

- For Amazon EMR release 6.9.0 and later, all components installed by Amazon EMR that use Log4j libraries use Log4j version 2.17.1 or later.
- When you use the DynamoDB connector with Spark on Amazon EMR versions 6.6.0, 6.7.0, and 6.8.0, all reads from your table return an empty result, even though the input split references non-empty data. Amazon EMR release 6.9.0 fixes this issue.
- Amazon EMR 6.9.0 adds limited support for Lake Formation-based access control with Apache Hudi when reading data using Spark SQL. The support is for SELECT queries using Spark SQL and is limited to column-level access control. For more information, see Hudi and Lake Formation.
- When you use Amazon EMR 6.9.0 to create a Hadoop cluster with Node Labels enabled, the YARN metrics API returns aggregated information across all partitions, instead of the default partition. For more information, see YARN-11414.
- With Amazon EMR release 6.9.0, we've updated Trino to version 398, which uses Java 17. The previous supported version of Trino for Amazon EMR 6.8.0 was Trino 388 running on Java 11. For more information about this change, see Trino updates to Java 17 on the Trino blog.
- This releases fixes a timing sequence mismatch issue between Apache BigTop and the Amazon EMR on EC2 cluster startup sequence. This timing sequence mismatch occurs when a system attempts to perform two or more operations at the same time instead of doing them in the proper sequence. As a result, certain cluster configurations experienced instance startup timeouts and slower cluster startup times.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

> **Note**
>
> This release no longer gets automatic AMI updates since it has been succeeded by 1 more more patch releases. The patch release is denoted by the number after the second decimal point (`6.8.`*`1`*). To see if you're using the latest patch release, check the available releases in the *Release Guide*, or check the **Amazon EMR release** dropdown when you create a cluster in the console, or use the `ListReleaseLabels` API or `list-release-labels` CLI action. To get updates about new releases, subscribe to the RSS feed on the What's new? page.

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240816.0	4.14.350	August 21, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024809.0	4.14.349	August 20, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240719.0	4.14.348	July 25, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230808.0	4.14.320	August 24, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230727.0	4.14.320	August 14, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 719.0	4.14.320	August 2, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 628.0	4.14.318	July 12, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230612.0	4.14.314	June 23, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 504.1	4.14.313	May 16, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230418.0	4.14.311	May 3, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazoı Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 404.1	4.14.311	April 18, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)
2.0.2023 404.0	4.14.311	April 10, 2023	US East (N. Virginia), Europe (Paris)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230320.0	4.14.309	March 30, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 307.0	4.14.305	March 15, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 207.0	4.14.304	February 22, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2022 210.1	4.14.301	January 12, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2022 103.3	4.14.296	December 5, 2022	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

6.9.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.1.0	Delta lake is an open table format for huge analytic datasets
emr-ddb	4.16.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.3.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.6.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.7.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.23.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.2.0	EMR S3Select Connector
emrfs	2.54.0	Amazon S3 connector for Hadoop ecosystem applicati ons.

Component	Version	Description
flink-client	1.15.2	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.15.2	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	3.3.3-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.3.3-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	3.3.3-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	3.3.3-amzn-1	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	3.3.3-amzn-1	HDFS service for managing the Hadoop filesystem journal on HA clusters.

Component	Version	Description
hadoop-httpfs-server	3.3.3-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.3.3-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.3.3-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.3.3-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.3.3-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.3.3-amzn-1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	2.4.13-amzn-0	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.4.13-amzn-0	Service for serving one or more HBase regions.
hbase-client	2.4.13-amzn-0	HBase command-line client.
hbase-rest-server	2.4.13-amzn-0	Service providing a RESTful HTTP endpoint for HBase.

Component	Version	Description
hbase-thrift-server	2.4.13-amzn-0	Service providing a Thrift endpoint to HBase.
hbase-operator-tools	2.4.13-amzn-0	Repair tool for Apache HBase clusters.
hcatalog-client	3.1.3-amzn-2	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.3-amzn-2	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.3-amzn-2	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.3-amzn-2	Hive command line client.
hive-hbase	3.1.3-amzn-2	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-2	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	3.1.3-amzn-2	Service for accepting Hive queries as web requests.
hudi	0.12.1-amzn-0	Incremental processing framework to power data pipeline at low latency and high efficiency.

Component	Version	Description
hudi-presto	0.12.1-amzn-0	Bundle library for running Presto with Hudi.
hudi-trino	0.12.1-amzn-0	Bundle library for running Trino with Hudi.
hudi-spark	0.12.1-amzn-0	Bundle library for running Spark with Hudi.
hue-server	4.10.0	Web application for analyzing data using Hadoop ecosystem applications
iceberg	0.14.1-amzn-0	Apache Iceberg is an open table format for huge analytic datasets
jupyterhub	1.4.1	Multi-user server for Jupyter notebooks
livy-server	0.7.1-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.9.1	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68+	MariaDB database server.
nvidia-cuda	11.7.0	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.

Component	Version	Description
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	The phoenix libraries for server and client
phoenix-connectors	6.0.0-SNAPSHOT	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	6.0.0	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.276-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.276-amzn-0	Service for executing pieces of a query.
presto-client	0.276-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
trino-coordinator	398-amzn-0	Service for accepting queries and managing query execution among trino-wor kers.

Component	Version	Description
trino-worker	398-amzn-0	Service for executing pieces of a query.
trino-client	398-amzn-0	Trino command-line client which is installed on an HA cluster's stand-by masters where Trino server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.3.0-amzn-1	Spark command-line clients.
spark-history-server	3.3.0-amzn-1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	3.3.0-amzn-1	In-memory execution engine for YARN.
spark-yarn-slave	3.3.0-amzn-1	Apache Spark libraries needed by YARN slaves.
spark-rapids	22.08.0-amzn-0	Nvidia Spark RAPIDS plugin that accelerates Apache Spark with GPUs.
sqoop-client	1.4.7	Apache Sqoop command-line client.

Component	Version	Description
tensorflow	2.10.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.10.2-amzn-0	The tez YARN application and libraries.
webserver	2.4.41+	Apache HTTP server.
zeppelin-server	0.10.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.5.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.5.10	ZooKeeper command line client.

6.9.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-6.9.0 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN

Classifications	Description	Reconfiguration Actions
		services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-ses sion.properties settings for Kubernetes/Yarn session.	Restarts Flink history server.
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Restarts Flink history server.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts PhoenixQu eryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classifications	Description	Reconfiguration Actions
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	This classification should not be reconfigured.
hdfs-env	Change values in the HDFS environment.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat server.

Classifications	Description	Reconfiguration Actions
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat server.
hive	Amazon EMR-curated settings for Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.
hive-env	Change values in the Hive environment.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Not available.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.

Classifications	Description	Reconfiguration Actions
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Change values in the Hudi environment.	Not available.
hudi-defaults	Change values in Hudi's hudi-defaults.conf file.	Not available.
iceberg-defaults	Change values in Iceberg's iceberg-defaults.conf file.	Not available.

Classifications	Description	Reconfiguration Actions
delta-defaults	Change values in Delta's delta-defaults.conf file.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j2	Change Livy log4j2.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.

Classifications	Description	Reconfiguration Actions
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hb ase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServe r.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.pro perties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoDB)
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoDB)

Classifications	Description	Reconfiguration Actions
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoDB)
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-lakeformation	Change values in Presto's lakeformation.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
trino-log	Change values in Trino's log.properties file.	Restarts Trino-Server (for Trino)
trino-config	Change values in Trino's config.properties file.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Change values in Trino's password-authenticator.prop erties file.	Restarts Trino-Server (for Trino)
trino-env	Change values in Trino's trino-env.sh file.	Restarts Trino-Server (for Trino)
trino-node	Change values in Trino's node.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
trino-connector-blackhole	Change values in Trino's blackhole.properties file.	Not available.
trino-connector-cassandra	Change values in Trino's cassandra.properties file.	Not available.
trino-connector-delta	Change values in Trino's delta.properties file.	Restarts Trino-Server (for Trino)
trino-connector-hive	Change values in Trino's hive.properties file.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Change values in Trino's exchange-manager.properties file.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Change values in Trino's iceberg.properties file.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Change values in Trino's jmx.properties file.	Not available.
trino-connector-kafka	Change values in Trino's kafka.properties file.	Not available.
trino-connector-localfile	Change values in Trino's localfile.properties file.	Not available.
trino-connector-memory	Change values in Trino's memory.properties file.	Not available.
trino-connector-mongodb	Change values in Trino's mongodb.properties file.	Not available.
trino-connector-mysql	Change values in Trino's mysql.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
trino-connector-postgresql	Change values in Trino's postgresql.properties file.	Not available.
trino-connector-raptor	Change values in Trino's raptor.properties file.	Not available.
trino-connector-redis	Change values in Trino's redis.properties file.	Not available.
trino-connector-redshift	Change values in Trino's redshift.properties file.	Not available.
trino-connector-tpch	Change values in Trino's tpch.properties file.	Not available.
trino-connector-tpcds	Change values in Trino's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-log4j	Change values in kms-log4j .properties file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.

Classifications	Description	Reconfiguration Actions
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j2	Change values in Spark's log4j2.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restart Oozie and HiveServe r2.
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts MapReduce -HistoryServer.

Classifications	Description	Reconfiguration Actions
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Livy Server and MapReduce-HistoryS erver.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

6.9.0 change log

Change log for 6.9.0 release and release notes

Date	Event	Description
2023-08-30	Update release notes	Added fix for timing sequence mismatch issue
2023-08-21	Update release notes	Added a known issue with Hadoop 3.3.3.
2023-07-26	Update	New OS release labels `2.0.20230612.0` and `2.0.20230628.0` .

Date	Event	Description
2022-12-13	Release notes updated	Added feature and known issue for runtime with SageMaker
2022-11-29	Release notes and documenta tion updated	Added feature for Amazon Redshift integration for Apache Spark
2022-11-23	Release notes updated	Removed Log4j entry
2022-11-18	Deployment complete	Amazon EMR 6.9 fully deployed to all supported Regions
2022-11-18	Docs publication	Amazon EMR 6.9 release notes first published
2022-11-14	Initial release	Amazon EMR 6.9 deployed to limited commercial Regions

Amazon EMR release 6.8.1

6.8.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, Iceberg, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Trino, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases

- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.8.1	emr-6.8.0	emr-6.7.0	emr-6.6.0
AWS SDK for Java	1.12.170	1.12.170	1.12.170	1.12.170
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.12.15	2.12.15	2.12.15	2.12.10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.15.1	1.15.1	1.14.2	1.14.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.12	2.4.12	2.4.4	2.4.4
HCatalog	3.1.3	3.1.3	3.1.3	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.3	3.1.3	3.1.3	3.1.2
Hudi	0.11.1-amzn-0	0.11.1-amzn-0	0.11.0-amzn-0	0.10.1-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	0.14.0-amzn-0	0.14.0-amzn-0	0.13.1-amzn-0	0.13.1
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1

	emr-6.8.1	emr-6.8.0	emr-6.7.0	emr-6.6.0
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.8.0	1.8.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.273	0.273	0.272	0.267
Spark	3.3.0	3.3.0	3.2.1	3.2.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.9.1	2.9.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	388	388	378	367
Zeppelin	0.10.1	0.10.1	0.10.0	0.10.0
ZooKeeper	3.5.10	3.5.10	3.5.7	3.5.7

6.8.1 release notes

The following release notes include information for Amazon EMR release 6.8.1. Changes are relative to 6.8.0. For information on the release timeline, see the 6.8.1 change log.

Changes, enhancements, and resolved issues

- Hadoop 3.3.3 introduced a change in YARN (YARN-9608) that keeps nodes where containers ran in a decommissioning state until the application completes. This change ensures that local data such as shuffle data doesn't get lost, and you don' need to re-run the job. This approach might also lead to underutilization of resources on clusters with or without managed scaling enabled.

With Amazon EMR releases 6.11.0 and higher as well as 6.8.1, 6.9.1, and 6.10.1, the value of `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` is set to `false` in `yarn-site.xml` to resolve this issue.

While the fix addresses the issues that were introduced by YARN-9608, it might cause Hive jobs to fail due to shuffle data loss on clusters that have managed scaling enabled. We've mitigated that risk in this release by also setting `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` for Hive workloads. This config is only available with Amazon EMR releases 6.11.0 and higher.

- Metrics collector won't send any metrics to the control plane after failover of primary node in clusters with the instance groups configuration.
- This release eliminates retries on failed HTTP requests to metrics collector endpoints.
- This release includes a change that allows high-availability clusters to recover from failed state after restart.
- This release fixes an issue where large user-created UIDs caused overflow exceptions.
- This release fixes timeout issues with the Amazon EMR reconfiguration process.
- This release prevents an issue where failed reconfiguration might break other, unrelated processes.
- This release includes security fixes.
- This release fixes an issue where clusters that are running workloads on Spark with Amazon EMR might silently receive incorrect results with `contains`, `startsWith`, `endsWith`, and `like`. This issue occurs when you use the expressions on partitioned fields that have metadata in the Amazon EMR Hive3 Metastore Server (HMS).
- With Amazon EMR 6.6.0 through 6.9.x, INSERT queries with dynamic partition and an ORDER BY or SORT BY clause will always have two reducers. This issue is caused by OSS change HIVE-20703, which puts dynamic sort partition optimization under cost-based decision. If your workload doesn't require sorting of dynamic partitions, we recommend that you set the `hive.optimize.sort.dynamic.partition.threshold` property to `-1` to disable the new feature and get the correctly calculated number of reducers. This issue is fixed in OSS Hive as part of HIVE-22269 and is fixed in Amazon EMR 6.10.0.
- Hive might experience data loss when you use HDFS as a scratch directory and you have enabled merge small files, and the table contains static partition paths.

- This release fixes a performance issue with Hive if merge small files (disabled by default) is enabled at the end of ETL job.
- This release fixes an issue with throttling on the Glue side when there are no user-defined functions (UDF).
- This release fixes an issue that deletes container logs by the node log aggregation service before log pusher can push them to S3 in case of YARN decommissioning.
- This release fixes handling of compacted/archived files with persistent storefile tracking for HBase.
- This release fixes an issue that impacted Spark performance when you set a default `true` value for the `spark.yarn.heterogeneousExecutors.enabled` config in `spark-defaults.conf`.
- This release fixes an issue with Reduce Task failing to read shuffle data. The issue caused Hive query failures with a corrupted memory error.
- This release fixes an issue that caused the node provisioner to fail if the HDFS NameNode (NN) service was stuck in safemode during node replacement.
- This release adds a new retry mechanism to the cluster scaling workflow for EMR clusters that run Presto or Trino. This improvement reduces the risk that cluster resizing will indefinitely stall due to a single failed resize operation. It also improves cluster utilization, because your cluster scales up and down faster.
- This release improves cluster scale-down logic so that your cluster doesn't attempt a scale-down of core nodes below the HDFS replication factor setting for the cluster. This aligns with your data redundancy requirements, and reduces the chance that a scaling operation might stall.
- The log management daemon has been upgraded to identify all logs that are in active use with open file handles on the local instance storage, and the associated processes. This upgrade ensures that Amazon EMR properly deletes the files and reclaims storage space after the logs are archived to Amazon S3.
- This release includes a log-management daemon enhancement that deletes empty, unused steps directories in the local cluster file system. An excessively large number of empty directories can degrade the performance of Amazon EMR daemons and result in disk over-utilization.
- This release fixes an issue that might occur when you create an edge node by replicating one of the primary nodes from a cluster with multiple primary nodes. The replicated edge node could cause delays with scale-down operations, or result in high memory-utilization on the primary nodes. For more information on how to create an edge node to communicate with your EMR cluster, see [Edge Node Creator](#) in the `aws-samples` repo on GitHub.

- This release improves the automation process that Amazon EMR uses to re-mount Amazon EBS volumes to an instance after a reboot.
- This release fixes an issue that resulted in intermittent gaps in the Hadoop metrics that Amazon EMR publishes to Amazon CloudWatch.
- This release fixes an issue with EMR clusters where an update to the YARN configuration file that contains the exclusion list of nodes for the cluster is interrupted due to disk over-utilization. The incomplete update hinders future cluster scale-down operations. This release ensures that your cluster remains healthy, and that scaling operations work as expected.
- This release improves the on-cluster log management daemon to monitor additional log folders in your EMR cluster. This improvement minimizes disk over-utilization scenarios.
- This release automatically restarts the on-cluster log management daemon when it stops. This improvement reduces the risk for nodes to appear unhealthy due to disk over-utilization.
- This release adds support for archiving logs to Amazon S3 during cluster scale-down. Previously, you could only archive log files to Amazon S3 during cluster termination. The new capability ensures that log files generated on the cluster persist on Amazon S3 even after the node is terminated. For more information, see Configure cluster logging and debugging.
- This release fixes an issue that occurred when the Amazon S3 URI for a bootstrap action ended with a port number, for example: `a.b.c.d:4345`. Amazon EMR was incorrectly parsing these URIs, so any associated bootstrap actions would fail.
- This releases fixes a timing sequence mismatch issue between Apache BigTop and the Amazon EMR on EC2 cluster startup sequence. This timing sequence mismatch occurs when a system attempts to perform two or more operations at the same time instead of doing them in the proper sequence. As a result, certain cluster configurations experienced instance startup timeouts and slower cluster startup times.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 816.0	4.14.350	August 21, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240809.0	4.14.349	August 20, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 719.0	4.14.348	July 25, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv),

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 223.0	4.14.336	March 8, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240131.0	4.14.336	February 14, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 124.0	4.14.336	February 7, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West) , AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240109.0	4.14.334	January 24, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 218.0	4.14.330	January 2, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20231206.0	4.14.330	December 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20231116.0	4.14.328	December 11, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20231101.0	4.14.327	November 16, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20231020.1	4.14.326	November 7, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 012.1	4.14.326	October 26, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West) , AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230926.0	4.14.322	October 19, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230906.0	4.14.322	October 4, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230822.0	4.14.322	Augest 30, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023[illegible]808.0	4.14.320	August 24, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Canada (Central)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 727.0	4.14.320	August 14, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

6.8.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.2.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.7.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.22.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.1.0	EMR S3Select Connector
emrfs	2.53.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.15.1	Apache Flink command line client scripts and applications.

Component	Version	Description
flink-jobmanager-config	1.15.1	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	3.2.1-amzn-8.1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-8.1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	3.2.1-amzn-8.1	HDFS command-line client and library
hadoop-hdfs-namenode	3.2.1-amzn-8.1	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	3.2.1-amzn-8.1	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	3.2.1-amzn-8.1	HTTP endpoint for HDFS operations.

Component	Version	Description
hadoop-kms-server	3.2.1-amzn-8.1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.2.1-amzn-8.1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-8.1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.2.1-amzn-8.1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.2.1-amzn-8.1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	2.4.12-amzn-0.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.4.12-amzn-0.1	Service for serving one or more HBase regions.
hbase-client	2.4.12-amzn-0.1	HBase command-line client.
hbase-rest-server	2.4.12-amzn-0.1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	2.4.12-amzn-0.1	Service providing a Thrift endpoint to HBase.

Component	Version	Description
hbase-operator-tools	2.4.12-amzn-0.1	Repair tool for Apache HBase clusters.
hcatalog-client	3.1.3-amzn-1.1	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.3-amzn-1.1	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.3-amzn-1.1	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.3-amzn-1.1	Hive command line client.
hive-hbase	3.1.3-amzn-1.1	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-1.1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	3.1.3-amzn-1.1	Service for accepting Hive queries as web requests.
hudi	0.11.1-amzn-0	Incremental processing framework to power data pipeline at low latency and high efficiency.
hudi-presto	0.11.1-amzn-0	Bundle library for running Presto with Hudi.

Component	Version	Description
hudi-trino	0.11.1-amzn-0	Bundle library for running Trino with Hudi.
hudi-spark	0.11.1-amzn-0	Bundle library for running Spark with Hudi.
hue-server	4.10.0	Web application for analyzing data using Hadoop ecosystem applications
iceberg	0.14.0-amzn-0	Apache Iceberg is an open table format for huge analytic datasets
jupyterhub	1.4.1	Multi-user server for Jupyter notebooks
livy-server	0.7.1-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.9.1	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68+	MariaDB database server.
nvidia-cuda	11.7.0	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.
oozie-server	5.2.1	Service for accepting Oozie workflow requests.

Component	Version	Description
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	The phoenix libraries for server and client
phoenix-connectors	5.1.2	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.273.3-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.273.3-amzn-0	Service for executing pieces of a query.
presto-client	0.273.3-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
trino-coordinator	388-amzn-0	Service for accepting queries and managing query execution among trino-wor kers.
trino-worker	388-amzn-0	Service for executing pieces of a query.

Component	Version	Description
trino-client	388-amzn-0	Trino command-line client which is installed on an HA cluster's stand-by masters where Trino server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.3.0-amzn-0.1	Spark command-line clients.
spark-history-server	3.3.0-amzn-0.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	3.3.0-amzn-0.1	In-memory execution engine for YARN.
spark-yarn-slave	3.3.0-amzn-0.1	Apache Spark libraries needed by YARN slaves.
spark-rapids	22.06.0-amzn-0	Nvidia Spark RAPIDS plugin that accelerates Apache Spark with GPUs.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.9.1	TensorFlow open source software library for high performance numerical computation.

Component	Version	Description
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.41+	Apache HTTP server.
zeppelin-server	0.10.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.5.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.5.10	ZooKeeper command line client.

6.8.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-6.8.1 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.

Classifications	Description	Reconfiguration Actions
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.

Classifications	Description	Reconfiguration Actions
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-session.properties settings for Kubernetes/Yarn session.	Restarts Flink history server.
flink-log4j-cli	Change Flink log4j-cli.properties settings.	Restarts Flink history server.

Classifications	Description	Reconfiguration Actions
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts PhoenixQu eryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.

Classifications	Description	Reconfiguration Actions
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	This classification should not be reconfigured.
hdfs-env	Change values in the HDFS environment.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classifications	Description	Reconfiguration Actions
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat server.
hive	Amazon EMR-curated settings for Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
hive-env	Change values in the Hive environment.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Not available.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.

Classifications	Description	Reconfiguration Actions
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Change values in the Hudi environment.	Not available.
hudi-defaults	Change values in Hudi's hudi-defaults.conf file.	Not available.
iceberg-defaults	Change values in Iceberg's iceberg-defaults.conf file.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.

Classifications	Description	Reconfiguration Actions
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j	Change Livy log4j.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hb ase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServe r.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.pro perties file.	Not available.

Classifications	Description	Reconfiguration Actions
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoDB)
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoDB)
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
trino-log	Change values in Trino's log.properties file.	Restarts Trino-Server (for Trino)
trino-config	Change values in Trino's config.properties file.	Restarts Trino-Server (for Trino)

Classifications	Description	Reconfiguration Actions
trino-password-authenticator	Change values in Trino's password-authenticator.properties file.	Restarts Trino-Server (for Trino)
trino-env	Change values in Trino's trino-env.sh file.	Restarts Trino-Server (for Trino)
trino-node	Change values in Trino's node.properties file.	Not available.
trino-connector-blackhole	Change values in Trino's blackhole.properties file.	Not available.
trino-connector-cassandra	Change values in Trino's cassandra.properties file.	Not available.
trino-connector-hive	Change values in Trino's hive.properties file.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Change values in Trino's iceberg.properties file.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Change values in Trino's jmx.properties file.	Not available.
trino-connector-kafka	Change values in Trino's kafka.properties file.	Not available.
trino-connector-localfile	Change values in Trino's localfile.properties file.	Not available.
trino-connector-memory	Change values in Trino's memory.properties file.	Not available.
trino-connector-mongodb	Change values in Trino's mongodb.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
trino-connector-mysql	Change values in Trino's mysql.properties file.	Not available.
trino-connector-postgresql	Change values in Trino's postgresql.properties file.	Not available.
trino-connector-raptor	Change values in Trino's raptor.properties file.	Not available.
trino-connector-redis	Change values in Trino's redis.properties file.	Not available.
trino-connector-redshift	Change values in Trino's redshift.properties file.	Not available.
trino-connector-tpch	Change values in Trino's tpch.properties file.	Not available.
trino-connector-tpcds	Change values in Trino's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-log4j	Change values in kms-log4j .properties file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.

Classifications	Description	Reconfiguration Actions
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j2	Change values in Spark's log4j2.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restart Oozie and HiveServe r2.
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts MapReduce -HistoryServer.

Classifications	Description	Reconfiguration Actions
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Livy Server and MapReduce-HistoryS erver.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

6.8.1 change log

Change log for 6.8.1 release and release notes

Date	Event	Description
2023-08-30	Update release notes	Added several control-plane related fixes to the release notes
2023-08-21	Docs publication	Amazon EMR 6.8.1 release notes first published
2023-08-16	Deployment complete	Amazon EMR 6.8.1 fully deployed to all supported Regions

Date	Event	Description
2023-08-04	Initial release	Amazon EMR 6.8.1 first deployed to limited commercial Regions

Amazon EMR release 6.8.0

6.8.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, Iceberg, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Trino, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.8.0	emr-6.7.0	emr-6.6.0	emr-6.5.0
AWS SDK for Java	1.12.170	1.12.170	1.12.170	1.12.31
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.12.15	2.12.15	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-

	emr-6.8.0	emr-6.7.0	emr-6.6.0	emr-6.5.0
Delta	-	-	-	-
Flink	1.15.1	1.14.2	1.14.2	1.14.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.12	2.4.4	2.4.4	2.4.4
HCatalog	3.1.3	3.1.3	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.3	3.1.3	3.1.2	3.1.2
Hudi	0.11.1-amzn-0	0.11.0-amzn-0	0.10.1-amzn-0	0.9.0-amzn-1
Hue	4.10.0	4.10.0	4.10.0	4.9.0
Iceberg	0.14.0-amzn-0	0.13.1-amzn-0	0.13.1	0.12.0
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.8.0	1.8.0	1.8.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.273	0.272	0.267	0.261
Spark	3.3.0	3.2.1	3.2.0	3.1.2

	emr-6.8.0	emr-6.7.0	emr-6.6.0	emr-6.5.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.9.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	388	378	367	360
Zeppelin	0.10.1	0.10.0	0.10.0	0.10.0
ZooKeeper	3.5.10	3.5.7	3.5.7	3.5.7

6.8.0 release notes

The following release notes include information for Amazon EMR release 6.8.0. Changes are relative to 6.7.0.

New Features

- Amazon EMR steps feature now supports Apache Livy endpoint and JDBC/ODBC clients. For more information, see Configure runtime roles for Amazon EMR steps.
- Amazon EMR release 6.8.0 comes with Apache HBase release 2.4.12. With this HBase release, you can both archive and delete your HBase tables. The Amazon S3 archive process renames all table files to the archive directory. This can be a costly and lengthy process. Now, you can skip the archive process and quickly drop and delete large tables. For more information, see Using the HBase shell.

Known Issues

- Hadoop 3.3.3 introduced a change in YARN (YARN-9608) that keeps nodes where containers ran in a decommissioning state until the application completes. This change ensures that local data such as shuffle data doesn't get lost, and you don' need to re-run the job. In Amazon EMR 6.8.0 and 6.9.0, this approach might also lead to underutilization of resources on clusters with or without managed scaling enabled.

With Amazon EMR 6.10.0, there's a workaround for this issue to set the value of `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` to `false` in `yarn-site.xml`. In Amazon EMR releases 6.11.0 and higher as well as 6.8.1, 6.9.1, and 6.10.1, the config is set to `false` by default to resolve this issue.

Changes, Enhancements, and Resolved Issues

- When Amazon EMR release 6.5.0, 6.6.0, or 6.7.0 read Apache Phoenix tables through the Apache Spark shell, Amazon EMR produced a `NoSuchMethodError`. Amazon EMR release 6.8.0 fixes this issue.
- Amazon EMR release 6.8.0 comes with Apache Hudi 0.11.1; however, Amazon EMR 6.8.0 clusters are also compatible with the open-source `hudi-spark3.3-bundle_2.12` from Hudi 0.12.0.
- Amazon EMR release 6.8.0 comes with Apache Spark 3.3.0. This Spark release uses Apache Log4j 2 and the `log4j2.properties` file to configure Log4j in Spark processes. If you use Spark in the cluster or create EMR clusters with custom configuration parameters, and you want to upgrade to Amazon EMR release 6.8.0, you must migrate to the new `spark-log4j2` configuration classification and key format for Apache Log4j 2. For more information, see Migrating from Apache Log4j 1.x to Log4j 2.x.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

> **Note**
>
> This release no longer gets automatic AMI updates since it has been succeeded by 1 more more patch releases. The patch release is denoted by the number after the second decimal point (`6.8.1`). To see if you're using the latest patch release, check the available releases in the *Release Guide*, or check the **Amazon EMR release** dropdown when you create a cluster in the console, or use the `ListReleaseLabels` API or `list-release-labels` CLI action. To get updates about new releases, subscribe to the RSS feed on the What's new? page.

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 816.0	4.14.350	August 21, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 809.0	4.14.349	August 20, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240719.0	4.14.348	July 25, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 808.0	4.14.320	August 24, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230727.0	4.14.320	August 14, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central),

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 719.0	4.14.320	August 2, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 628.0	4.14.318	July 12, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 612.0	4.14.314	June 23, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230504.1	4.14.313	May 16, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 418.0	4.14.311	May 3, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230404.1	4.14.311	April 18, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)
2.0.20230404.0	4.14.311	April 10, 2023	US East (N. Virginia), Europe (Paris)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230320.0	4.14.309	March 30, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 307.0	4.14.305	March 15, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 207.0	4.14.304	February 22, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 119.1	4.14.301	February 3, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 210.1	4.14.301	December 22, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 103.3	4.14.296	December 5, 2022	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20221004.0	4.14.294	November 2, 2022	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelease Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20220912.1	4.14.291	September 6, 2022	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

Known Issues

- When you use the DynamoDB connector with Spark on Amazon EMR versions 6.6.0, 6.7.0, and 6.8.0, all reads from your table return an empty result, even though the input split references non-empty data. This is because Spark 3.2.0 sets `spark.hadoopRDD.ignoreEmptySplits` to `true` by default. As a workaround, explicitly set `spark.hadoopRDD.ignoreEmptySplits` to `false`. Amazon EMR release 6.9.0 fixes this issue.

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&‘()*+,-`. For more information, see [UTF-8 encoding table and Unicode characters](#).

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.
- With Amazon EMR releases 5.36.0 and 6.6.0 through 6.9.0, `SecretAgent` and `RecordServer` service components may experience log data loss due to an incorrect file name pattern configuration in Log4j2 properties. The incorrect configuration causes the components to generate only one log file per day. When the rotation strategy occurs, it overwrites the existing file instead of generating a new log file as expected. As a workaround, use a bootstrap action to generate log files each hour and append an auto-increment integer in the file name to handle the rotation.

  For Amazon EMR 6.6.0 through 6.9.0 releases, use the following bootstrap action when you launch a cluster.

  ```
  --bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
  ```

  For Amazon EMR 5.36.0, use the following bootstrap action when you launch a cluster.

  ```
  --bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
  ```

For more information on the release timeline, see the [change log](#).

6.8.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.2.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.7.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.22.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.1.0	EMR S3Select Connector

Component	Version	Description
emrfs	2.53.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.15.1	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.15.1	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	3.2.1-amzn-8	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-8	HDFS node-level service for storing blocks.
hadoop-hdfs-library	3.2.1-amzn-8	HDFS command-line client and library
hadoop-hdfs-namenode	3.2.1-amzn-8	HDFS service for tracking file names and block locations.

Component	Version	Description
hadoop-hdfs-journalnode	3.2.1-amzn-8	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	3.2.1-amzn-8	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.2.1-amzn-8	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.2.1-amzn-8	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-8	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.2.1-amzn-8	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.2.1-amzn-8	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	2.4.12-amzn-0	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.4.12-amzn-0	Service for serving one or more HBase regions.
hbase-client	2.4.12-amzn-0	HBase command-line client.

Component	Version	Description
hbase-rest-server	2.4.12-amzn-0	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	2.4.12-amzn-0	Service providing a Thrift endpoint to HBase.
hbase-operator-tools	2.4.12-amzn-0	Repair tool for Apache HBase clusters.
hcatalog-client	3.1.3-amzn-1	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.3-amzn-1	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.3-amzn-1	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.3-amzn-1	Hive command line client.
hive-hbase	3.1.3-amzn-1	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	3.1.3-amzn-1	Service for accepting Hive queries as web requests.

Component	Version	Description
hudi	0.11.1-amzn-0	Incremental processing framework to power data pipeline at low latency and high efficiency.
hudi-presto	0.11.1-amzn-0	Bundle library for running Presto with Hudi.
hudi-trino	0.11.1-amzn-0	Bundle library for running Trino with Hudi.
hudi-spark	0.11.1-amzn-0	Bundle library for running Spark with Hudi.
hue-server	4.10.0	Web application for analyzing data using Hadoop ecosystem applications
iceberg	0.14.0-amzn-0	Apache Iceberg is an open table format for huge analytic datasets
jupyterhub	1.4.1	Multi-user server for Jupyter notebooks
livy-server	0.7.1-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.9.1	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68+	MariaDB database server.

Component	Version	Description
nvidia-cuda	11.7.0	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	The phoenix libraries for server and client
phoenix-connectors	5.1.2	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.273.3-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.273.3-amzn-0	Service for executing pieces of a query.
presto-client	0.273.3-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.

Component	Version	Description
trino-coordinator	388-amzn-0	Service for accepting queries and managing query execution among trino-wor kers.
trino-worker	388-amzn-0	Service for executing pieces of a query.
trino-client	388-amzn-0	Trino command-line client which is installed on an HA cluster's stand-by masters where Trino server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.3.0-amzn-0	Spark command-line clients.
spark-history-server	3.3.0-amzn-0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	3.3.0-amzn-0	In-memory execution engine for YARN.
spark-yarn-slave	3.3.0-amzn-0	Apache Spark libraries needed by YARN slaves.
spark-rapids	22.06.0-amzn-0	Nvidia Spark RAPIDS plugin that accelerates Apache Spark with GPUs.

Component	Version	Description
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.9.1	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.41+	Apache HTTP server.
zeppelin-server	0.10.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.5.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.5.10	ZooKeeper command line client.

6.8.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-6.8.0 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN

Classifications	Description	Reconfiguration Actions
		services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-ses sion.properties settings for Kubernetes/Yarn session.	Restarts Flink history server.
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Restarts Flink history server.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts PhoenixQu eryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classifications	Description	Reconfiguration Actions
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	This classification should not be reconfigured.
hdfs-env	Change values in the HDFS environment.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat server.

Classifications	Description	Reconfiguration Actions
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat server.
hive	Amazon EMR-curated settings for Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.
hive-env	Change values in the Hive environment.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Not available.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.

Classifications	Description	Reconfiguration Actions
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Change values in the Hudi environment.	Not available.
hudi-defaults	Change values in Hudi's hudi-defaults.conf file.	Not available.
iceberg-defaults	Change values in Iceberg's iceberg-defaults.conf file.	Not available.

Classifications	Description	Reconfiguration Actions
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j	Change Livy log4j.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.

Classifications	Description	Reconfiguration Actions
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hb ase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServe r.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.pro perties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoDB)
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticat or	Change values in Presto's password-authenticator.prop erties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoDB)
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
trino-log	Change values in Trino's log.properties file.	Restarts Trino-Server (for Trino)
trino-config	Change values in Trino's config.properties file.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Change values in Trino's password-authenticator.prop erties file.	Restarts Trino-Server (for Trino)
trino-env	Change values in Trino's trino-env.sh file.	Restarts Trino-Server (for Trino)
trino-node	Change values in Trino's node.properties file.	Not available.
trino-connector-blackhole	Change values in Trino's blackhole.properties file.	Not available.
trino-connector-cassandra	Change values in Trino's cassandra.properties file.	Not available.
trino-connector-hive	Change values in Trino's hive.properties file.	Restarts Trino-Server (for Trino)

Classifications	Description	Reconfiguration Actions
trino-connector-iceberg	Change values in Trino's iceberg.properties file.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Change values in Trino's jmx.properties file.	Not available.
trino-connector-kafka	Change values in Trino's kafka.properties file.	Not available.
trino-connector-localfile	Change values in Trino's localfile.properties file.	Not available.
trino-connector-memory	Change values in Trino's memory.properties file.	Not available.
trino-connector-mongodb	Change values in Trino's mongodb.properties file.	Not available.
trino-connector-mysql	Change values in Trino's mysql.properties file.	Not available.
trino-connector-postgresql	Change values in Trino's postgresql.properties file.	Not available.
trino-connector-raptor	Change values in Trino's raptor.properties file.	Not available.
trino-connector-redis	Change values in Trino's redis.properties file.	Not available.
trino-connector-redshift	Change values in Trino's redshift.properties file.	Not available.
trino-connector-tpch	Change values in Trino's tpch.properties file.	Not available.
trino-connector-tpcds	Change values in Trino's tpcds.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-log4j	Change values in kms-log4j .properties file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j2	Change values in Spark's log4j2.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.

Classifications	Description	Reconfiguration Actions
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restart Oozie and HiveServe r2.
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts MapReduce -HistoryServer.
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Livy Server and MapReduce-HistoryS erver.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

6.8.0 change log

Change log for 6.8.0 release and release notes

Date	Event	Description
2023-08-21	Update	Added a known issue with Hadoop 3.3.3.
2023-07-26	Update	New OS release labels `2.0.20230612.0` and `2.0.20230628.0` .
2022-09-06	Deployment complete	Amazon EMR 6.8 fully deployed to all supported Regions
2022-09-06	Initial publication	Amazon EMR 6.8 release notes first published
2022-08-31	Initial release	Amazon EMR 6.8 released to limited commercial Regions

Amazon EMR release 6.7.0

6.7.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, Iceberg, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Trino, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases

- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.7.0	emr-6.6.0	emr-6.5.0	emr-6.4.0
AWS SDK for Java	1.12.170	1.12.170	1.12.31	1.12.31
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.12.15	2.12.10	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.14.2	1.14.0	1.13.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.4	2.4.4	2.4.4	2.4.4
HCatalog	3.1.3	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.3	3.1.2	3.1.2	3.1.2
Hudi	0.11.0-amzn-0	0.10.1-amzn-0	0.9.0-amzn-1	0.8.0-amzn-0
Hue	4.10.0	4.10.0	4.9.0	4.9.0
Iceberg	0.13.1-amzn-0	0.13.1	0.12.0	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1

	emr-6.7.0	emr-6.6.0	emr-6.5.0	emr-6.4.0
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.8.0	1.8.0	1.8.0	1.8.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.272	0.267	0.261	0.254.1
Spark	3.2.1	3.2.0	3.1.2	3.1.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	378	367	360	359
Zeppelin	0.10.0	0.10.0	0.10.0	0.9.0
ZooKeeper	3.5.7	3.5.7	3.5.7	3.5.7

6.7.0 release notes

The following release notes include information for Amazon EMR release 6.7.0. Changes are relative to 6.6.0.

Initial release date: July 15, 2022

New Features

- Amazon EMR now supports Apache Spark 3.2.1, Apache Hive 3.1.3, HUDI 0.11, PrestoDB 0.272, and Trino 0.378.
- Supports IAM Role and Lake Formation-based access controls with EMR steps (Spark, Hive) for Amazon EMR on EC2 clusters.
- Supports Apache Spark data definition statements on Apache Ranger enabled clusters. This now includes support for Trino applications reading and writing Apache Hive metadata on Apache Ranger enabled clusters. For more information, see Enable federated governance using Trino and Apache Ranger on Amazon EMR.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240816.0	4.14.350	August 21, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
			(Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 809.0	4.14.349	August 20, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 719.0	4.14.348	July 25, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 709.1	4.14.348	July 23, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing) , China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240223.0	4.14.336	March 8, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240131.0	4.14.336	February 14, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazoı Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 124.0	4.14.336	February 7, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240109.0	4.14.334	January 24, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelease Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 218.0	4.14.330	January 2, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20231206.0	4.14.330	December 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 116.0	4.14.328	December 11, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazoı Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 101.0	4.14.327	November 16, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 020.1	4.14.326	November 7, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20231012.1	4.14.326	October 26, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230926.0	4.14.322	October 19, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 906.0	4.14.322	October 4, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

OsRelea Label (Amazoı Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 822.0	4.14.322	August 30, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230808.0	4.14.320	August 24, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 727.0	4.14.320	August 14, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230719.0	4.14.320	August 2, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 628.0	4.14.318	July 12, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 612.0	4.14.314	June 23, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230504.1	4.14.313	May 16, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 418.0	4.14.311	May 3, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 404.1	4.14.311	April 18, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)
2.0.2023 404.0	4.14.311	April 10, 2023	US East (N. Virginia), Europe (Paris)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230320.0	4.14.309	March 30, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230307.0	4.14.305	March 15, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazoı Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 207.0	4.14.304	February 22, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 119.1	4.14.301	February 3, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazoı Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 210.1	4.14.301	December 22, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20221103.3	4.14.296	December 5, 2022	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 004.0	4.14.294	November 2, 2022	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022912.1	4.14.291	October 7, 2022	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)
2.0.2022719.0	4.14.287	August 10, 2022	`us-west-1` , `eu-west-3` , `eu-north-1` , `ap-south-1` , `me-south-1`

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 606.1	4.14.281	July 15, 2022	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

Known Issues

- When Amazon EMR release 6.5.0, 6.6.0, or 6.7.0 read Apache Phoenix tables through the Apache Spark shell, a `NoSuchMethodError` occurs because Amazon EMR uses an incorrect `Hbase.compat.version`. Amazon EMR release 6.8.0 fixes this issue.
- When you use the DynamoDB connector with Spark on Amazon EMR versions 6.6.0, 6.7.0, and 6.8.0, all reads from your table return an empty result, even though the input split references non-empty data. This is because Spark 3.2.0 sets `spark.hadoopRDD.ignoreEmptySplits` to

`true` by default. As a workaround, explicitly set `spark.hadoopRDD.ignoreEmptySplits` to `false`. Amazon EMR release 6.9.0 fixes this issue.

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&‘()*+,-`. For more information, see UTF-8 encoding table and Unicode characters.

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

- With Amazon EMR releases 5.36.0 and 6.6.0 through 6.9.0, `SecretAgent` and `RecordServer` service components may experience log data loss due to an incorrect file name pattern configuration in Log4j2 properties. The incorrect configuration causes the components to generate only one log file per day. When the rotation strategy occurs, it overwrites the existing file instead of generating a new log file as expected. As a workaround, use a bootstrap action to generate log files each hour and append an auto-increment integer in the file name to handle the rotation.

  For Amazon EMR 6.6.0 through 6.9.0 releases, use the following bootstrap action when you launch a cluster.

  ```
  --bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
  ```

  For Amazon EMR 5.36.0, use the following bootstrap action when you launch a cluster.

  ```
  --bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
  ```

- The `GetClusterSessionCredentials` API isn't supported with clusters that run on Amazon EMR 6.7 or lower.

6.7.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.2.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.6.0	Conda env for emr notebook which includes jupyter enterprise gateway

Component	Version	Description
emr-s3-dist-cp	2.22.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.1.0	EMR S3Select Connector
emrfs	2.52.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.14.2	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.14.2	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	3.2.1-amzn-7	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-7	HDFS node-level service for storing blocks.

Component	Version	Description
hadoop-hdfs-library	3.2.1-amzn-7	HDFS command-line client and library
hadoop-hdfs-namenode	3.2.1-amzn-7	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	3.2.1-amzn-7	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	3.2.1-amzn-7	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.2.1-amzn-7	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.2.1-amzn-7	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-7	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.2.1-amzn-7	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.2.1-amzn-7	Service for retrieving current and historical information for YARN applications.

Component	Version	Description
hbase-hmaster	2.4.4-amzn-3	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.4.4-amzn-3	Service for serving one or more HBase regions.
hbase-client	2.4.4-amzn-3	HBase command-line client.
hbase-rest-server	2.4.4-amzn-3	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	2.4.4-amzn-3	Service providing a Thrift endpoint to HBase.
hbase-operator-tools	2.4.4-amzn-3	Repair tool for Apache HBase clusters.
hcatalog-client	3.1.3-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.3-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.3-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.3-amzn-0	Hive command line client.
hive-hbase	3.1.3-amzn-0	Hive-hbase client.

Component	Version	Description
hive-metastore-server	3.1.3-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	3.1.3-amzn-0	Service for accepting Hive queries as web requests.
hudi	0.11.0-amzn-0	Incremental processing framework to power data pipeline at low latency and high efficiency.
hudi-presto	0.11.0-amzn-0	Bundle library for running Presto with Hudi.
hudi-trino	0.11.0-amzn-0	Bundle library for running Trino with Hudi.
hudi-spark	0.11.0-amzn-0	Bundle library for running Spark with Hudi.
hue-server	4.10.0	Web application for analyzing data using Hadoop ecosystem applications
iceberg	0.13.1-amzn-0	Apache Iceberg is an open table format for huge analytic datasets
jupyterhub	1.4.1	Multi-user server for Jupyter notebooks
livy-server	0.7.1-incubating	REST interface for interacting with Apache Spark

Component	Version	Description
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.8.0	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68+	MariaDB database server.
nvidia-cuda	11.0.194	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	The phoenix libraries for server and client
phoenix-connectors	5.1.2	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.272-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.

Component	Version	Description
presto-worker	0.272-amzn-0	Service for executing pieces of a query.
presto-client	0.272-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
trino-coordinator	378-amzn-0	Service for accepting queries and managing query execution among trino-wor kers.
trino-worker	378-amzn-0	Service for executing pieces of a query.
trino-client	378-amzn-0	Trino command-line client which is installed on an HA cluster's stand-by masters where Trino server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.2.1-amzn-0	Spark command-line clients.
spark-history-server	3.2.1-amzn-0	Web UI for viewing logged events for the lifetime of a completed Spark application.

Component	Version	Description
spark-on-yarn	3.2.1-amzn-0	In-memory execution engine for YARN.
spark-yarn-slave	3.2.1-amzn-0	Apache Spark libraries needed by YARN slaves.
spark-rapids	22.02.0-amzn-1	Nvidia Spark RAPIDS plugin that accelerates Apache Spark with GPUs.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.4.1	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.41+	Apache HTTP server.
zeppelin-server	0.10.0	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.5.7	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.5.7	ZooKeeper command line client.

6.7.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-6.7.0 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-ses sion.properties settings for Kubernetes/Yarn session.	Restarts Flink history server.
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Restarts Flink history server.

Classifications	Description	Reconfiguration Actions
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services SecondaryNamenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.

Classifications	Description	Reconfiguration Actions
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	This classification should not be reconfigured.
hdfs-env	Change values in the HDFS environment.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classifications	Description	Reconfiguration Actions
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat server.
hive	Amazon EMR-curated settings for Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
hive-env	Change values in the Hive environment.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Not available.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.

Classifications	Description	Reconfiguration Actions
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Change values in the Hudi environment.	Not available.
hudi-defaults	Change values in Hudi's hudi-defaults.conf file.	Not available.
iceberg-defaults	Change values in Iceberg's iceberg-defaults.conf file.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.

Classifications	Description	Reconfiguration Actions
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j	Change Livy log4j.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hb ase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServe r.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.pro perties file.	Not available.

Classifications	Description	Reconfiguration Actions
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoDB)
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoDB)
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
trino-log	Change values in Trino's log.properties file.	Restarts Trino-Server (for Trino)
trino-config	Change values in Trino's config.properties file.	Restarts Trino-Server (for Trino)

Classifications	Description	Reconfiguration Actions
trino-password-authenticator	Change values in Trino's password-authenticator.properties file.	Restarts Trino-Server (for Trino)
trino-env	Change values in Trino's trino-env.sh file.	Restarts Trino-Server (for Trino)
trino-node	Change values in Trino's node.properties file.	Not available.
trino-connector-blackhole	Change values in Trino's blackhole.properties file.	Not available.
trino-connector-cassandra	Change values in Trino's cassandra.properties file.	Not available.
trino-connector-hive	Change values in Trino's hive.properties file.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Change values in Trino's iceberg.properties file.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Change values in Trino's jmx.properties file.	Not available.
trino-connector-kafka	Change values in Trino's kafka.properties file.	Not available.
trino-connector-localfile	Change values in Trino's localfile.properties file.	Not available.
trino-connector-memory	Change values in Trino's memory.properties file.	Not available.
trino-connector-mongodb	Change values in Trino's mongodb.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
trino-connector-mysql	Change values in Trino's mysql.properties file.	Not available.
trino-connector-postgresql	Change values in Trino's postgresql.properties file.	Not available.
trino-connector-raptor	Change values in Trino's raptor.properties file.	Not available.
trino-connector-redis	Change values in Trino's redis.properties file.	Not available.
trino-connector-redshift	Change values in Trino's redshift.properties file.	Not available.
trino-connector-tpch	Change values in Trino's tpch.properties file.	Not available.
trino-connector-tpcds	Change values in Trino's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-log4j	Change values in kms-log4j .properties file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.

Classifications	Description	Reconfiguration Actions
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j	Change values in Spark's log4j.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restart Oozie and HiveServer2.
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Livy Server and MapReduce-HistoryS erver.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

Amazon EMR release 6.6.0

6.6.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, Iceberg, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Trino, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases

- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.6.0	emr-6.5.0	emr-6.4.0	emr-6.3.1
AWS SDK for Java	1.12.170	1.12.31	1.12.31	1.11.977
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.12.10	2.12.10	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.14.0	1.13.1	1.12.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.4	2.4.4	2.4.4	2.2.6
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.10.1-amzn-0	0.9.0-amzn-1	0.8.0-amzn-0	0.7.0-amzn-0
Hue	4.10.0	4.9.0	4.9.0	4.9.0
Iceberg	0.13.1	0.12.0	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.2.2

	emr-6.6.0	emr-6.5.0	emr-6.4.0	emr-6.3.1
Livy	0.7.1	0.7.1	0.7.1	0.7.0
MXNet	1.8.0	1.8.0	1.8.0	1.7.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.267	0.261	0.254.1	0.245.1
Spark	3.2.0	3.1.2	3.1.2	3.1.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	367	360	359	350
Zeppelin	0.10.0	0.10.0	0.9.0	0.9.0
ZooKeeper	3.5.7	3.5.7	3.5.7	3.4.14

6.6.0 release notes

The following release notes include information for Amazon EMR release 6.6.0. Changes are relative to 6.5.0.

Initial release date: May 9, 2022

Updated documentation date: June 15, 2022

New Features

- Amazon EMR 6.6 now supports Apache Spark 3.2, Apache Spark RAPIDS 22.02, CUDA 11, Apache Hudi 0.10.1, Apache Iceberg 0.13, Trino 0.367 and PrestoDB 0.267.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see [Using the default Amazon Linux AMI for Amazon EMR](#).

OsReleaseLabel (Amazon Linux Version)	**Amazon Linux Kernel Version**	**Available Date**	**Supported Regions**
2.0.20240816.0	4.14.350	August 21, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
			(US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)
2.0.20240809.0	4.14.349	August 20, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 719.0	4.14.348	July 25, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 223.0	4.14.336	March 8, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 131.0	4.14.336	February 14, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 124.0	4.14.336	February 7, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 109.0	4.14.334	January 24, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 218.0	4.14.330	January 2, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelease Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 206.0	4.14.330	December 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 116.0	4.14.328	December 11, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20231101.0	4.14.327	November 16, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20231020.1	4.14.326	November 7, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20231012.1	4.14.326	October 26, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 926.0	4.14.322	October 19, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelease Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 906.0	4.14.322	October 4, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 822.0	4.14.322	August 30, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

OsRelease Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 808.0	4.14.320	August 24, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 727.0	4.14.320	August 14, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230719.0	4.14.320	August 2, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023...628.0	4.14.318	July 12, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 612.0	4.14.314	June 23, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 504.1	4.14.313	May 16, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230418.0	4.14.311	May 3, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 404.1	4.14.311	April 18, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)
2.0.2023 404.0	4.14.311	April 10, 2023	US East (N. Virginia), Europe (Paris)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230320.0	4.14.309	March 30, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 307.0	4.14.305	March 15, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230207.0	4.14.304	February 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230119.1	4.14.301	February 3, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 210.1	4.14.301	December 22, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20221103.3	4.14.296	December 5, 2022	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20221004.0	4.14.294	November 2, 2022	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 912.1	4.14.291	October 7, 2022	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)
2.0.2022 805.0	4.14.287	August 30, 2022	`us-west-1`

OsRelease Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 719.0	4.14.287	August 10, 2022	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 426.0	4.14.281	June 10, 2022	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 406.1	4.14.275	May 2, 2022	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

- With Amazon EMR 6.6 and later, applications that use Log4j 1.x and Log4j 2.x are upgraded to use Log4j 1.2.17 (or higher) and Log4j 2.17.1 (or higher) respectively, and do not require using the bootstrap actions provided to mitigate the CVE issues.
- **[Managed scaling] Spark shuffle data managed scaling optimization** - For Amazon EMR versions 5.34.0 and later, and EMR versions 6.4.0 and later, managed scaling is now Spark shuffle data aware (data that Spark redistributes across partitions to perform specific operations). For more information on shuffle operations, see Using EMR managed scaling in Amazon EMR in the *Amazon EMR Management Guide* and Spark Programming Guide.

- Starting with Amazon EMR 5.32.0 and 6.5.0, dynamic executor sizing for Apache Spark is enabled by default. To turn this feature on or off, you can use the `spark.yarn.heterogeneousExecutors.enabled` configuration parameter.

Changes, Enhancements, and Resolved Issues

- Amazon EMR reduces cluster startup time by up to 80 seconds on average for clusters that use the EMR default AMI option and only install common applications, such as Apache Hadoop, Apache Spark and Apache Hive.

Known Issues

- When Amazon EMR release 6.5.0, 6.6.0, or 6.7.0 read Apache Phoenix tables through the Apache Spark shell, a `NoSuchMethodError` occurs because Amazon EMR uses an incorrect `Hbase.compat.version`. Amazon EMR release 6.8.0 fixes this issue.
- When you use the DynamoDB connector with Spark on Amazon EMR versions 6.6.0, 6.7.0, and 6.8.0, all reads from your table return an empty result, even though the input split references non-empty data. This is because Spark 3.2.0 sets `spark.hadoopRDD.ignoreEmptySplits` to `true` by default. As a workaround, explicitly set `spark.hadoopRDD.ignoreEmptySplits` to `false`. Amazon EMR release 6.9.0 fixes this issue.
- On Trino long-running clusters, Amazon EMR 6.6.0 enables Garbage Collection logging parameters in the Trino jvm.config to get better insights from the Garbage Collection logs. This change appends many Garbage Collection logs to the launcher.log (/var/log/trino/launcher.log) file. If you are running Trino clusters in Amazon EMR 6.6.0, you may encounter nodes running out of disk space after the cluster has been running for a couple of days due to the appended logs.

  The workaround for this issue is to run the script below as a Bootstrap Action to disable the Garbage Collection logging parameters in jvm.config while creating or cloning the cluster for Amazon EMR 6.6.0.

  ```
  #!/bin/bash
    set -ex
    PRESTO_PUPPET_DIR='/var/aws/emr/bigtop-deploy/puppet/modules/trino'
    sudo bash -c "sed -i '/-Xlog/d' ${PRESTO_PUPPET_DIR}/templates/jvm.config"
  ```

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:

  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&‘()*+,-`. For more information, see [UTF-8 encoding table and Unicode characters](#).

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

- With Amazon EMR releases 5.36.0 and 6.6.0 through 6.9.0, `SecretAgent` and `RecordServer` service components may experience log data loss due to an incorrect file name pattern configuration in Log4j2 properties. The incorrect configuration causes the components to generate only one log file per day. When the rotation strategy occurs, it overwrites the existing file instead of generating a new log file as expected. As a workaround, use a bootstrap action to generate log files each hour and append an auto-increment integer in the file name to handle the rotation.

  For Amazon EMR 6.6.0 through 6.9.0 releases, use the following bootstrap action when you launch a cluster.

  ```
  --bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
  ```

  For Amazon EMR 5.36.0, use the following bootstrap action when you launch a cluster.

  ```
  --bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
  ```

6.6.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.2.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.5.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.20.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.1.0	EMR S3Select Connector

Component	Version	Description
emrfs	2.50.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.14.2	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.14.2	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	3.2.1-amzn-6	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-6	HDFS node-level service for storing blocks.
hadoop-hdfs-library	3.2.1-amzn-6	HDFS command-line client and library
hadoop-hdfs-namenode	3.2.1-amzn-6	HDFS service for tracking file names and block locations.

Component	Version	Description
hadoop-hdfs-journalnode	3.2.1-amzn-6	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	3.2.1-amzn-6	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.2.1-amzn-6	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.2.1-amzn-6	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-6	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.2.1-amzn-6	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.2.1-amzn-6	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	2.4.4-amzn-2	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.4.4-amzn-2	Service for serving one or more HBase regions.
hbase-client	2.4.4-amzn-2	HBase command-line client.

Component	Version	Description
hbase-rest-server	2.4.4-amzn-2	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	2.4.4-amzn-2	Service providing a Thrift endpoint to HBase.
hbase-operator-tools	2.4.4-amzn-2	Repair tool for Apache HBase clusters.
hcatalog-client	3.1.2-amzn-7	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.2-amzn-7	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.2-amzn-7	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.2-amzn-7	Hive command line client.
hive-hbase	3.1.2-amzn-7	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-7	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	3.1.2-amzn-7	Service for accepting Hive queries as web requests.

Component	Version	Description
hudi	0.10.1-amzn-0	Incremental processing framework to power data pipeline at low latency and high efficiency.
hudi-presto	0.10.1-amzn-0	Bundle library for running Presto with Hudi.
hudi-trino	0.10.1-amzn-0	Bundle library for running Trino with Hudi.
hudi-spark	0.10.1-amzn-0	Bundle library for running Spark with Hudi.
hue-server	4.10.0	Web application for analyzing data using Hadoop ecosystem applications
iceberg	0.13.1	Apache Iceberg is an open table format for huge analytic datasets
jupyterhub	1.4.1	Multi-user server for Jupyter notebooks
livy-server	0.7.1-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.8.0	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68+	MariaDB database server.

Component	Version	Description
nvidia-cuda	11.0.194	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	The phoenix libraries for server and client
phoenix-connectors	5.1.2	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.267-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.267-amzn-0	Service for executing pieces of a query.
presto-client	0.267-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.

Component	Version	Description
trino-coordinator	367-amzn-0	Service for accepting queries and managing query execution among trino-wor kers.
trino-worker	367-amzn-0	Service for executing pieces of a query.
trino-client	367-amzn-0	Trino command-line client which is installed on an HA cluster's stand-by masters where Trino server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.2.0-amzn-0	Spark command-line clients.
spark-history-server	3.2.0-amzn-0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	3.2.0-amzn-0	In-memory execution engine for YARN.
spark-yarn-slave	3.2.0-amzn-0	Apache Spark libraries needed by YARN slaves.
spark-rapids	22.02.0-amzn-0	Nvidia Spark RAPIDS plugin that accelerates Apache Spark with GPUs.

Component	Version	Description
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.4.1	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.41+	Apache HTTP server.
zeppelin-server	0.10.0	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.5.7	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.5.7	ZooKeeper command line client.

6.6.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-6.6.0 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN

Classifications	Description	Reconfiguration Actions
		services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-ses sion.properties settings for Kubernetes/Yarn session.	Restarts Flink history server.
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Restarts Flink history server.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts PhoenixQu eryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classifications	Description	Reconfiguration Actions
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	This classification should not be reconfigured.
hdfs-env	Change values in the HDFS environment.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat server.

Classifications	Description	Reconfiguration Actions
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat server.
hive	Amazon EMR-curated settings for Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.
hive-env	Change values in the Hive environment.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Not available.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.

Classifications	Description	Reconfiguration Actions
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Change values in the Hudi environment.	Not available.
hudi-defaults	Change values in Hudi's hudi-defaults.conf file.	Not available.
iceberg-defaults	Change values in Iceberg's iceberg-defaults.conf file.	Not available.

Classifications	Description	Reconfiguration Actions
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j	Change Livy log4j.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.

Classifications	Description	Reconfiguration Actions
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hb ase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServe r.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.pro perties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoDB)
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticat or	Change values in Presto's password-authenticator.prop erties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoDB)
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
trino-log	Change values in Trino's log.properties file.	Restarts Trino-Server (for Trino)
trino-config	Change values in Trino's config.properties file.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Change values in Trino's password-authenticator.prop erties file.	Restarts Trino-Server (for Trino)
trino-env	Change values in Trino's trino-env.sh file.	Restarts Trino-Server (for Trino)
trino-node	Change values in Trino's node.properties file.	Not available.
trino-connector-blackhole	Change values in Trino's blackhole.properties file.	Not available.
trino-connector-cassandra	Change values in Trino's cassandra.properties file.	Not available.
trino-connector-hive	Change values in Trino's hive.properties file.	Restarts Trino-Server (for Trino)

Classifications	Description	Reconfiguration Actions
trino-connector-iceberg	Change values in Trino's iceberg.properties file.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Change values in Trino's jmx.properties file.	Not available.
trino-connector-kafka	Change values in Trino's kafka.properties file.	Not available.
trino-connector-localfile	Change values in Trino's localfile.properties file.	Not available.
trino-connector-memory	Change values in Trino's memory.properties file.	Not available.
trino-connector-mongodb	Change values in Trino's mongodb.properties file.	Not available.
trino-connector-mysql	Change values in Trino's mysql.properties file.	Not available.
trino-connector-postgresql	Change values in Trino's postgresql.properties file.	Not available.
trino-connector-raptor	Change values in Trino's raptor.properties file.	Not available.
trino-connector-redis	Change values in Trino's redis.properties file.	Not available.
trino-connector-redshift	Change values in Trino's redshift.properties file.	Not available.
trino-connector-tpch	Change values in Trino's tpch.properties file.	Not available.
trino-connector-tpcds	Change values in Trino's tpcds.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-log4j	Change values in kms-log4j .properties file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j	Change values in Spark's log4j.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.

Classifications	Description	Reconfiguration Actions
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restart Oozie and HiveServer2.
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

Amazon EMR release 6.5.0

6.5.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, Iceberg, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Trino, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.5.0	emr-6.4.0	emr-6.3.1	emr-6.3.0
AWS SDK for Java	1.12.31	1.12.31	1.11.977	1.11.977
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.12.10	2.12.10	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.0	1.13.1	1.12.1	1.12.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.4	2.4.4	2.2.6	2.2.6

	emr-6.5.0	emr-6.4.0	emr-6.3.1	emr-6.3.0
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.9.0-amzn-1	0.8.0-amzn-0	0.7.0-amzn-0	0.7.0-amzn-0
Hue	4.9.0	4.9.0	4.9.0	4.9.0
Iceberg	0.12.0	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.2.2	1.2.2
Livy	0.7.1	0.7.1	0.7.0	0.7.0
MXNet	1.8.0	1.8.0	1.7.0	1.7.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.261	0.254.1	0.245.1	0.245.1
Spark	3.1.2	3.1.2	3.1.1	3.1.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-6.5.0	emr-6.4.0	emr-6.3.1	emr-6.3.0
Trino (PrestoSQL)	360	359	350	350
Zeppelin	0.10.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.5.7	3.5.7	3.4.14	3.4.14

6.5.0 release notes

The following release notes include information for Amazon EMR release 6.5.0. Changes are relative to 6.4.0.

Initial release date: January 20, 2022

Updated release date: March 21, 2022

New Features

- **[Managed scaling] Spark shuffle data managed scaling optimization** - For Amazon EMR versions 5.34.0 and later, and EMR versions 6.4.0 and later, managed scaling is now Spark shuffle data aware (data that Spark redistributes across partitions to perform specific operations). For more information on shuffle operations, see Using EMR managed scaling in Amazon EMR in the *Amazon EMR Management Guide* and Spark Programming Guide.
- Starting with Amazon EMR 5.32.0 and 6.5.0, dynamic executor sizing for Apache Spark is enabled by default. To turn this feature on or off, you can use the `spark.yarn.heterogeneousExecutors.enabled` configuration parameter.
- Support for Apache Iceberg open table format for huge analytic datasets.
- Support for ranger-trino-plugin 2.0.1-amzn-1
- Support for toree 0.5.0

Changes, Enhancements, and Resolved Issues

- Amazon EMR 6.5 release version now supports Apache Iceberg 0.12.0, and provides runtime improvements with Amazon EMR Runtime for Apache Spark, Amazon EMR Runtime for Presto, and Amazon EMR Runtime for Apache Hive.

- Apache Iceberg is an open table format for large data sets in Amazon S3 and provides fast query performance over large tables, atomic commits, concurrent writes, and SQL-compatible table evolution. With EMR 6.5, you can use Apache Spark 3.1.2 with the Iceberg table format.
- Apache Hudi 0.9 adds Spark SQL DDL and DML support. This allows you to create, upsert Hudi tables using just SQL statements. Apache Hudi 0.9 also includes query side and writer side performance improvements.
- Amazon EMR Runtime for Apache Hive improves Apache Hive performance on Amazon S3 by removing rename operations during staging operations, and improves performance for metastore check (MSCK) commands used for repairing tables.

Known Issues

- When Amazon EMR release 6.5.0, 6.6.0, or 6.7.0 read Apache Phoenix tables through the Apache Spark shell, a `NoSuchMethodError` occurs because Amazon EMR uses an incorrect `Hbase.compat.version`. Amazon EMR release 6.8.0 fixes this issue.
- Hbase bundle clusters in high availability (HA) fail to provision with the default volume size and instance type. The workaround for this issue is to increase the root volume size.
- To use Spark actions with Apache Oozie, you must add the following configuration to your Oozie `workflow.xml` file. Otherwise, several critical libraries such as Hadoop and EMRFS will be missing from the classpath of the Spark executors that Oozie launches.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&‘()*+,-`. For more information, see UTF-8 encoding table and Unicode characters.

As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

6.5.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.2.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.4.0	Conda env for emr notebook which includes jupyter enterprise gateway

Component	Version	Description
emr-s3-dist-cp	2.19.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.1.0	EMR S3Select Connector
emrfs	2.48.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.14.0	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.14.0	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	3.2.1-amzn-5	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-5	HDFS node-level service for storing blocks.

Component	Version	Description
hadoop-hdfs-library	3.2.1-amzn-5	HDFS command-line client and library
hadoop-hdfs-namenode	3.2.1-amzn-5	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	3.2.1-amzn-5	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	3.2.1-amzn-5	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.2.1-amzn-5	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.2.1-amzn-5	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-5	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.2.1-amzn-5	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.2.1-amzn-5	Service for retrieving current and historical information for YARN applications.

Component	Version	Description
hbase-hmaster	2.4.4-amzn-1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.4.4-amzn-1	Service for serving one or more HBase regions.
hbase-client	2.4.4-amzn-1	HBase command-line client.
hbase-rest-server	2.4.4-amzn-1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	2.4.4-amzn-1	Service providing a Thrift endpoint to HBase.
hcatalog-client	3.1.2-amzn-6	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.2-amzn-6	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.2-amzn-6	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.2-amzn-6	Hive command line client.
hive-hbase	3.1.2-amzn-6	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-6	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.

Component	Version	Description
hive-server2	3.1.2-amzn-6	Service for accepting Hive queries as web requests.
hudi	0.9.0-amzn-1	Incremental processing framework to power data pipeline at low latency and high efficiency.
hudi-presto	0.9.0-amzn-1	Bundle library for running Presto with Hudi.
hudi-trino	0.9.0-amzn-1	Bundle library for running Trino with Hudi.
hudi-spark	0.9.0-amzn-1	Bundle library for running Spark with Hudi.
hue-server	4.9.0	Web application for analyzing data using Hadoop ecosystem applications
iceberg	0.12.0	Apache Iceberg is an open table format for huge analytic datasets
jupyterhub	1.4.1	Multi-user server for Jupyter notebooks
livy-server	0.7.1-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.8.0	A flexible, scalable, and efficient library for deep learning.

Component	Version	Description
mariadb-server	5.5.68+	MariaDB database server.
nvidia-cuda	10.1.243	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	The phoenix libraries for server and client
phoenix-query-server	5.1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.261-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.261-amzn-0	Service for executing pieces of a query.
presto-client	0.261-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.

Component	Version	Description
trino-coordinator	360	Service for accepting queries and managing query execution among trino-wor kers.
trino-worker	360	Service for executing pieces of a query.
trino-client	360	Trino command-line client which is installed on an HA cluster's stand-by masters where Trino server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.1.2-amzn-1	Spark command-line clients.
spark-history-server	3.1.2-amzn-1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	3.1.2-amzn-1	In-memory execution engine for YARN.
spark-yarn-slave	3.1.2-amzn-1	Apache Spark libraries needed by YARN slaves.
spark-rapids	0.4.1	Nvidia Spark RAPIDS plugin that accelerates Apache Spark with GPUs.

Component	Version	Description
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.4.1	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.41+	Apache HTTP server.
zeppelin-server	0.10.0	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.5.7	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.5.7	ZooKeeper command line client.

6.5.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-6.5.0 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN

Classifications	Description	Reconfiguration Actions
		services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-ses sion.properties settings for Kubernetes/Yarn session.	Restarts Flink history server.
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Restarts Flink history server.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts PhoenixQu eryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classifications	Description	Reconfiguration Actions
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	This classification should not be reconfigured.
hdfs-env	Change values in the HDFS environment.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat server.

Classifications	Description	Reconfiguration Actions
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat server.
hive	Amazon EMR-curated settings for Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.
hive-env	Change values in the Hive environment.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Not available.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.

Classifications	Description	Reconfiguration Actions
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Change values in the Hudi environment.	Not available.
hudi-defaults	Change values in Hudi's hudi-defaults.conf file.	Not available.
iceberg-defaults	Change values in Iceberg's iceberg-defaults.conf file.	Not available.

Classifications	Description	Reconfiguration Actions
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j	Change Livy log4j.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.

Classifications	Description	Reconfiguration Actions
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hb ase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServe r.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.pro perties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoDB)
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticat or	Change values in Presto's password-authenticator.prop erties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoDB)
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
trino-log	Change values in Trino's log.properties file.	Restarts Trino-Server (for Trino)
trino-config	Change values in Trino's config.properties file.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Change values in Trino's password-authenticator.prop erties file.	Restarts Trino-Server (for Trino)
trino-env	Change values in Trino's trino-env.sh file.	Restarts Trino-Server (for Trino)
trino-node	Change values in Trino's node.properties file.	Not available.
trino-connector-blackhole	Change values in Trino's blackhole.properties file.	Not available.
trino-connector-cassandra	Change values in Trino's cassandra.properties file.	Not available.
trino-connector-hive	Change values in Trino's hive.properties file.	Restarts Trino-Server (for Trino)

Classifications	Description	Reconfiguration Actions
trino-connector-jmx	Change values in Trino's jmx.properties file.	Not available.
trino-connector-kafka	Change values in Trino's kafka.properties file.	Not available.
trino-connector-localfile	Change values in Trino's localfile.properties file.	Not available.
trino-connector-memory	Change values in Trino's memory.properties file.	Not available.
trino-connector-mongodb	Change values in Trino's mongodb.properties file.	Not available.
trino-connector-mysql	Change values in Trino's mysql.properties file.	Not available.
trino-connector-postgresql	Change values in Trino's postgresql.properties file.	Not available.
trino-connector-raptor	Change values in Trino's raptor.properties file.	Not available.
trino-connector-redis	Change values in Trino's redis.properties file.	Not available.
trino-connector-redshift	Change values in Trino's redshift.properties file.	Not available.
trino-connector-tpch	Change values in Trino's tpch.properties file.	Not available.
trino-connector-tpcds	Change values in Trino's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.

Classifications	Description	Reconfiguration Actions
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-log4j	Change values in kms-log4j .properties file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j	Change values in Spark's log4j.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.

Classifications	Description	Reconfiguration Actions
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restart Oozie and HiveServer2.
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

Amazon EMR release 6.4.0

6.4.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Trino, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.4.0	emr-6.3.1	emr-6.3.0	emr-6.2.1
AWS SDK for Java	1.12.31	1.11.977	1.11.977	1.11.880
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.12.10	2.12.10	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.13.1	1.12.1	1.12.1	1.11.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.4	2.2.6	2.2.6	2.2.6-amzn-0

	emr-6.4.0	emr-6.3.1	emr-6.3.0	emr-6.2.1
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.8.0-amzn-0	0.7.0-amzn-0	0.7.0-amzn-0	0.6.0-amzn-1
Hue	4.9.0	4.9.0	4.9.0	4.8.0
Iceberg	-	-	-	-
JupyterEn terpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.2.2	1.2.2	1.1.0
Livy	0.7.1	0.7.0	0.7.0	0.7.0
MXNet	1.8.0	1.7.0	1.7.0	1.7.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.0
Phoenix	5.1.2	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.254.1	0.245.1	0.245.1	0.238.3
Spark	3.1.2	3.1.1	3.1.1	3.0.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.3.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-6.4.0	emr-6.3.1	emr-6.3.0	emr-6.2.1
Trino (PrestoSQL)	359	350	350	343
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.5.7	3.4.14	3.4.14	3.4.14

6.4.0 release notes

The following release notes include information for Amazon EMR release 6.4.0. Changes are relative to 6.3.0.

Initial release date: Sept 20, 2021

Updated release date: March 21, 2022

Supported applications

- AWS SDK for Java version 1.12.31
- CloudWatch Sink version 2.2.0
- DynamoDB Connector version 4.16.0
- EMRFS version 2.47.0
- Amazon EMR Goodies version 3.2.0
- Amazon EMR Kinesis Connector version 3.5.0
- Amazon EMR Record Server version 2.1.0
- Amazon EMR Scripts version 2.5.0
- Flink version 1.13.1
- Ganglia version 3.7.2
- AWS Glue Hive Metastore Client version 3.3.0
- Hadoop version 3.2.1-amzn-4
- HBase version 2.4.4-amzn-0
- HBase-operator-tools 1.1.0
- HCatalog version 3.1.2-amzn-5

- Hive version 3.1.2-amzn-5
- Hudi version 0.8.0-amzn-0
- Hue version 4.9.0
- Java JDK version Corretto-8.302.08.1 (build 1.8.0_302-b08)
- JupyterHub version 1.4.1
- Livy version 0.7.1-incubating
- MXNet version 1.8.0
- Oozie version 5.2.1
- Phoenix version 5.1.2
- Pig version 0.17.0
- Presto version 0.254.1-amzn-0
- Trino version 359
- Apache Ranger KMS (multi-master transparent encryption) version 2.0.0
- ranger-plugins 2.0.1-amzn-0
- ranger-s3-plugin 1.2.0
- SageMaker Spark SDK version 1.4.1
- Scala version 2.12.10 (OpenJDK 64-Bit Server VM, Java 1.8.0_282)
- Spark version 3.1.2-amzn-0
- spark-rapids 0.4.1
- Sqoop version 1.4.7
- TensorFlow version 2.4.1
- tez version 0.9.2
- Zeppelin version 0.9.0
- Zookeeper version 3.5.7
- Connectors and drivers: DynamoDB Connector 4.16.0

New features

- **[Managed scaling] Spark shuffle data managed scaling optimization** - For Amazon EMR versions 5.34.0 and later, and EMR versions 6.4.0 and later, managed scaling is now Spark shuffle data aware (data that Spark redistributes across partitions to perform specific operations). For more information on shuffle operations, see Using EMR managed scaling in Amazon EMR in the *Amazon EMR Management Guide* and Spark Programming Guide.

- On Apache Ranger-enabled Amazon EMR clusters, you can use Apache Spark SQL to insert data into or update the Apache Hive metastore tables using `INSERT INTO`, `INSERT OVERWRITE`, and `ALTER TABLE`. When using ALTER TABLE with Spark SQL, a partition location must be the child directory of a table location. Amazon EMR does not currently support inserting data into a partition where the partition location is different from the table location.
- PrestoSQL has been renamed to Trino.
- Hive: Execution of simple SELECT queries with LIMIT clause are accelerated by stopping the query execution as soon as the number of records mentioned in LIMIT clause is fetched. Simple SELECT queries are queries that do not have GROUP BY / ORDER by clause or queries that do not have a reducer stage. For example, `SELECT * from <TABLE> WHERE <Condition> LIMIT <Number>`.

Hudi Concurrency Control

- Hudi now supports Optimistic Concurrency Control (OCC), which can be leveraged with write operations like UPSERT and INSERT to allow changes from multiple writers to the same Hudi table. This is file-level OCC, so any two commits (or writers) can write to the same table, if their changes do not conflict. For more information, see the Hudi concurrency control.
- Amazon EMR clusters have Zookeeper installed, which can be leveraged as the lock provider for OCC. To make it easier to use this feature, Amazon EMR clusters have the following properties pre-configured:

```
hoodie.write.lock.provider=org.apache.hudi.client.transaction.lock.ZookeeperBasedLockProvider
hoodie.write.lock.zookeeper.url=<EMR Zookeeper URL>
hoodie.write.lock.zookeeper.port=<EMR Zookeeper Port>
hoodie.write.lock.zookeeper.base_path=/hudi
```

To enable OCC, you need to configure the following properties either with their Hudi job options or at the cluster-level using the Amazon EMR configurations API:

```
hoodie.write.concurrency.mode=optimistic_concurrency_control
hoodie.cleaner.policy.failed.writes=LAZY (Performs cleaning of failed writes lazily
 instead of inline with every write)
hoodie.write.lock.zookeeper.lock_key=<Key to uniquely identify the Hudi table> (Table
 Name is a good option)
```

Hudi Monitoring: Amazon CloudWatch integration to report Hudi Metrics

- Amazon EMR supports publishing Hudi Metrics to Amazon CloudWatch. It is enabled by setting the following required configurations:

```
hoodie.metrics.on=true
hoodie.metrics.reporter.type=CLOUDWATCH
```

- The following are optional Hudi configurations that you can change:

Setting	Description	Value
hoodie.metrics.cloudwatch.report.period.seconds	Frequency (in seconds) at which to report metrics to Amazon CloudWatch	Default value is 60s, which is fine for the default one minute resolution offered by Amazon CloudWatch
hoodie.metrics.cloudwatch.metric.prefix	Prefix to be added to each metric name	Default value is empty (no prefix)
hoodie.metrics.cloudwatch.namespace	Amazon CloudWatch namespace under which metrics are published	Default value is Hudi
hoodie.metrics.cloudwatch.maxDatumsPerRequest	Maximum number of datums to be included in one request to Amazon CloudWatch	Default value is 20, which is same as Amazon CloudWatch default

Amazon EMR Hudi configurations support and improvements

- Customers can now leverage EMR Configurations API and Reconfiguration feature to configure Hudi configurations at cluster level. A new file based configuration support has been introduced via /etc/hudi/conf/hudi-defaults.conf along the lines of other applications like Spark, Hive etc. EMR configures few defaults to improve user experience:

  — `hoodie.datasource.hive_sync.jdbcurl` is configured to the cluster Hive server URL and no longer needs to be specified. This is particularly useful when running a job in Spark cluster mode, where you previously had to specify the Amazon EMR master IP.

— HBase specific configurations, which are useful for using HBase index with Hudi.

— Zookeeper lock provider specific configuration, as discussed under concurrency control, which makes it easier to use Optimistic Concurrency Control (OCC).

- Additional changes have been introduced to reduce the number of configurations that you need to pass, and to infer automatically where possible:

  — The `partitionBy` keyword can be used to specify the partition column.

  — When enabling Hive Sync, it is no longer mandatory to pass `HIVE_TABLE_OPT_KEY, HIVE_PARTITION_FIELDS_OPT_KEY, HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY`. Those values can be inferred from the Hudi table name and partition field.

  — `KEYGENERATOR_CLASS_OPT_KEY` is not mandatory to pass, and can be inferred from simpler cases of `SimpleKeyGenerator` and `ComplexKeyGenerator`.

Hudi Caveats

- Hudi does not support vectorized execution in Hive for Merge on Read (MoR) and Bootstrap tables. For example, `count(*)` fails with Hudi realtime table when `hive.vectorized.execution.enabled` is set to true. As a workaround, you can disable vectorized reading by setting `hive.vectorized.execution.enabled` to `false`.
- Multi-writer support is not compatible with the Hudi bootstrap feature.
- Flink Streamer and Flink SQL are experimental features in this release. These features are not recommended for use in production deployments.

Changes, enhancements, and resolved issues

This is a release to fix issues with Amazon EMR Scaling when it fails to scale up/scale down a cluster successfully or causes application failures.

- Previously, manual restart of the resource manager on a multi-master cluster caused Amazon EMR on-cluster daemons, like Zookeeper, to reload all previously decommissioned or lost nodes in the Zookeeper znode file. This caused default limits to be exceeded in certain situations. Amazon EMR now removes the decommissioned or lost node records older than one hour from the Zookeeper file and the internal limits have been increased.

- Fixed an issue where scaling requests failed for a large, highly utilized cluster when Amazon EMR on-cluster daemons were running health checking activities, such as gathering YARN node state and HDFS node state. This was happening because on-cluster daemons were not able to communicate the health status data of a node to internal Amazon EMR components.
- Improved EMR on-cluster daemons to correctly track the node states when IP addresses are reused to improve reliability during scaling operations.
- SPARK-29683. Fixed an issue where job failures occurred during cluster scale-down as Spark was assuming all available nodes were deny-listed.
- YARN-9011. Fixed an issue where job failures occurred due to a race condition in YARN decommissioning when cluster tried to scale up or down.
- Fixed issue with step or job failures during cluster scaling by ensuring that the node states are always consistent between the Amazon EMR on-cluster daemons and YARN/HDFS.
- Fixed an issue where cluster operations such as scale down and step submission failed for Amazon EMR clusters enabled with Kerberos authentication. This was because the Amazon EMR on-cluster daemon did not renew the Kerberos ticket, which is required to securely communicate with HDFS/YARN running on the primary node.
- **Configuring a cluster to fix Apache YARN Timeline Server version 1 and 1.5 performance issues**

  Apache YARN Timeline Server version 1 and 1.5 can cause performance issues with very active, large EMR clusters, particularly with `yarn.resourcemanager.system-metrics-publisher.enabled=true`, which is the default setting in Amazon EMR. An open source YARN Timeline Server v2 solves the performance issue related to YARN Timeline Server scalability.

  Other workarounds for this issue include:

  - Configuring yarn.resourcemanager.system-metrics-publisher.enabled=false in yarn-site.xml.
  - Enabling the fix for this issue when creating a cluster, as described below.

  The following Amazon EMR releases contain a fix for this YARN Timeline Server performance issue.

  EMR 5.30.2, 5.31.1, 5.32.1, 5.33.1, 5.34.x, 6.0.1, 6.1.1, 6.2.1, 6.3.1, 6.4.x

  To enable the fix on any of the above specified Amazon EMR releases, set these properties to `true` in a configurations JSON file that is passed in using the `aws emr create-cluster` command parameter: `--configurations file://./configurations.json`. Or enable the fix using the reconfiguration console UI.

Example of the configurations.json file contents:

```
[
{
"Classification": "yarn-site",
"Properties": {
"yarn.resourcemanager.system-metrics-publisher.timeline-server-v1.enable-batch":
 "true",
"yarn.resourcemanager.system-metrics-publisher.enabled": "true"
},
"Configurations": []
}
]
```

- WebHDFS and HttpFS server are disabled by default. You can re-enable WebHDFS using the Hadoop configuration, `dfs.webhdfs.enabled`. HttpFS server can be started by using `sudo systemctl start hadoop-httpfs`.
- HTTPS is now enabled by default for Amazon Linux repositories. If you are using an Amazon S3 VPCE policy to restrict access to specific buckets, you must add the new Amazon Linux bucket ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` to your policy (replace `$region` with the region where the endpoint is). For more information, see this topic in the AWS discussion forums. Announcement: Amazon Linux 2 now supports the ability to use HTTPS while connecting to package repositories .
- Hive: Write query performance is improved by enabling the use of a scratch directory on HDFS for the last job. The temporary data for final job is written to HDFS instead of Amazon S3 and performance is improved because the data is moved from HDFS to the final table location (Amazon S3) instead of between Amazon S3 devices.
- Hive: Query compilation time improvement up to 2.5x with Glue metastore Partition Pruning.
- By default, when built-in UDFs are passed by Hive to the Hive Metastore Server, only a subset of those built-in UDFs are passed to the Glue Metastore since Glue supports only limited expression operators. If you set `hive.glue.partition.pruning.client=true`, then all partition pruning happens on the client side. If the you set `hive.glue.partition.pruning.server=true`, then all partition pruning happens on the server side.

Known issues

- Hue queries do not work in Amazon EMR 6.4.0 because Apache Hadoop HttpFS server is disabled by default. To use Hue on Amazon EMR 6.4.0, either manually start HttpFS server on the Amazon EMR primary node using `sudo systemctl start hadoop-httpfs`, or use an Amazon EMR step.
- The Amazon EMR Notebooks feature used with Livy user impersonation does not work because HttpFS is disabled by default. In this case, the EMR notebook cannot connect to the cluster that has Livy impersonation enabled. The workaround is to start HttpFS server before connecting the EMR notebook to the cluster using `sudo systemctl start hadoop-httpfs`.
- In Amazon EMR version 6.4.0, Phoenix does not support the Phoenix connectors component.
- To use Spark actions with Apache Oozie, you must add the following configuration to your Oozie `workflow.xml` file. Otherwise, several critical libraries such as Hadoop and EMRFS will be missing from the classpath of the Spark executors that Oozie launches.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&‘()*+,-`. For more information, see UTF-8 encoding table and Unicode characters.

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

6.4.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.2.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.3.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.18.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.1.0	EMR S3Select Connector

Component	Version	Description
emrfs	2.47.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.13.1	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.13.1	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	3.2.1-amzn-4	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-4	HDFS node-level service for storing blocks.
hadoop-hdfs-library	3.2.1-amzn-4	HDFS command-line client and library
hadoop-hdfs-namenode	3.2.1-amzn-4	HDFS service for tracking file names and block locations.

Component	Version	Description
hadoop-hdfs-journalnode	3.2.1-amzn-4	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	3.2.1-amzn-4	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.2.1-amzn-4	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.2.1-amzn-4	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-4	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.2.1-amzn-4	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.2.1-amzn-4	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	2.4.4-amzn-0	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.4.4-amzn-0	Service for serving one or more HBase regions.
hbase-client	2.4.4-amzn-0	HBase command-line client.

Component	Version	Description
hbase-rest-server	2.4.4-amzn-0	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	2.4.4-amzn-0	Service providing a Thrift endpoint to HBase.
hcatalog-client	3.1.2-amzn-5	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.2-amzn-5	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.2-amzn-5	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.2-amzn-5	Hive command line client.
hive-hbase	3.1.2-amzn-5	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-5	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	3.1.2-amzn-5	Service for accepting Hive queries as web requests.
hudi	0.8.0-amzn-0	Incremental processing framework to power data pipline at low latency and high efficiency.

Component	Version	Description
hudi-presto	0.8.0-amzn-0	Bundle library for running Presto with Hudi.
hudi-trino	0.8.0-amzn-0	Bundle library for running Trino with Hudi.
hudi-spark	0.8.0-amzn-0	Bundle library for running Spark with Hudi.
hue-server	4.9.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.4.1	Multi-user server for Jupyter notebooks
livy-server	0.7.1-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.8.0	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68+	MariaDB database server.
nvidia-cuda	10.1.243	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.5.0	Open Source Computer Vision Library.

Component	Version	Description
phoenix-library	5.1.2	The phoenix libraries for server and client
phoenix-query-server	5.1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.254.1-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.254.1-amzn-0	Service for executing pieces of a query.
presto-client	0.254.1-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
trino-coordinator	359	Service for accepting queries and managing query execution among trino-wor kers.
trino-worker	359	Service for executing pieces of a query.
trino-client	359	Trino command-line client which is installed on an HA cluster's stand-by masters where Trino server is not started.

Component	Version	Description
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.1.2-amzn-0	Spark command-line clients.
spark-history-server	3.1.2-amzn-0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	3.1.2-amzn-0	In-memory execution engine for YARN.
spark-yarn-slave	3.1.2-amzn-0	Apache Spark libraries needed by YARN slaves.
spark-rapids	0.4.1	Nvidia Spark RAPIDS plugin that accelerates Apache Spark with GPUs.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.4.1	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.41+	Apache HTTP server.

Component	Version	Description
zeppelin-server	0.9.0	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.5.7	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.5.7	ZooKeeper command line client.

6.4.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-6.4.0 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.

Classifications	Description	Reconfiguration Actions
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.

Classifications	Description	Reconfiguration Actions
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-ses sion.properties settings for Kubernetes/Yarn session.	Restarts Flink history server.
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Restarts Flink history server.

Classifications	Description	Reconfiguration Actions
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services SecondaryNamenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.

Classifications	Description	Reconfiguration Actions
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hbase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	This classification should not be reconfigured.
hdfs-env	Change values in the HDFS environment.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classifications	Description	Reconfiguration Actions
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat server.
hive	Amazon EMR-curated settings for Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
hive-env	Change values in the Hive environment.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Not available.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.

Classifications	Description	Reconfiguration Actions
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Change values in the Hudi environment.	Not available.
hudi-defaults	Change values in Hudi's hudi-defaults.conf file.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.

Classifications	Description	Reconfiguration Actions
livy-log4j	Change Livy log4j.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServer.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.

Classifications	Description	Reconfiguration Actions
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoDB)
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoDB)
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
trino-log	Change values in Trino's log.properties file.	Restarts Trino-Server (for Trino)
trino-config	Change values in Trino's config.properties file.	Restarts Trino-Server (for Trino)

Classifications	Description	Reconfiguration Actions
trino-password-authenticator	Change values in Trino's password-authenticator.properties file.	Restarts Trino-Server (for Trino)
trino-env	Change values in Trino's trino-env.sh file.	Restarts Trino-Server (for Trino)
trino-node	Change values in Trino's node.properties file.	Not available.
trino-connector-blackhole	Change values in Trino's blackhole.properties file.	Not available.
trino-connector-cassandra	Change values in Trino's cassandra.properties file.	Not available.
trino-connector-hive	Change values in Trino's hive.properties file.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Change values in Trino's jmx.properties file.	Not available.
trino-connector-kafka	Change values in Trino's kafka.properties file.	Not available.
trino-connector-localfile	Change values in Trino's localfile.properties file.	Not available.
trino-connector-memory	Change values in Trino's memory.properties file.	Not available.
trino-connector-mongodb	Change values in Trino's mongodb.properties file.	Not available.
trino-connector-mysql	Change values in Trino's mysql.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
trino-connector-postgresql	Change values in Trino's postgresql.properties file.	Not available.
trino-connector-raptor	Change values in Trino's raptor.properties file.	Not available.
trino-connector-redis	Change values in Trino's redis.properties file.	Not available.
trino-connector-redshift	Change values in Trino's redshift.properties file.	Not available.
trino-connector-tpch	Change values in Trino's tpch.properties file.	Not available.
trino-connector-tpcds	Change values in Trino's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-log4j	Change values in kms-log4j .properties file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.

Classifications	Description	Reconfiguration Actions
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j	Change values in Spark's log4j.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restart Oozie and HiveServe r2.
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts MapReduce -HistoryServer.

Classifications	Description	Reconfiguration Actions
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Livy Server and MapReduce-HistoryS erver.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

Amazon EMR release 6.3.1

6.3.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, PrestoSQL, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases

- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.3.1	emr-6.3.0	emr-6.2.1	emr-6.2.0
AWS SDK for Java	1.11.977	1.11.977	1.11.880	1.11.880
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.12.10	2.12.10	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.12.1	1.12.1	1.11.2	1.11.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.6	2.2.6	2.2.6-amzn-0	2.2.6-amzn-0
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.7.0-amzn-0	0.7.0-amzn-0	0.6.0-amzn-1	0.6.0-amzn-1
Hue	4.9.0	4.9.0	4.8.0	4.8.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.2.2	1.2.2	1.1.0	1.1.0

	emr-6.3.1	emr-6.3.0	emr-6.2.1	emr-6.2.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.7.0	1.7.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.0	5.2.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.245.1	0.245.1	0.238.3	0.238.3
Spark	3.1.1	3.1.1	3.0.1	3.0.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.3.1	2.3.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	350	350	343	343
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

6.3.1 release notes

This is a release to fix issues with Amazon EMR Scaling when it fails to scale up/scale down a cluster successfully or causes application failures.

Changes, Enhancements, and Resolved Issues

- Fixed an issue where scaling requests failed for a large, highly utilized cluster when Amazon EMR on-cluster daemons were running health checking activities, such as gathering YARN node state and HDFS node state. This was happening because on-cluster daemons were not able to communicate the health status data of a node to internal Amazon EMR components.

- Improved EMR on-cluster daemons to correctly track the node states when IP addresses are reused to improve reliability during scaling operations.
- SPARK-29683. Fixed an issue where job failures occurred during cluster scale-down as Spark was assuming all available nodes were deny-listed.
- YARN-9011. Fixed an issue where job failures occurred due to a race condition in YARN decommissioning when cluster tried to scale up or down.
- Fixed issue with step or job failures during cluster scaling by ensuring that the node states are always consistent between the Amazon EMR on-cluster daemons and YARN/HDFS.
- Fixed an issue where cluster operations such as scale down and step submission failed for Amazon EMR clusters enabled with Kerberos authentication. This was because the Amazon EMR on-cluster daemon did not renew the Kerberos ticket, which is required to securely communicate with HDFS/YARN running on the primary node.
- HTTPS is now enabled by default for Amazon Linux repositories. If you are using an Amazon S3 VPCE policy to restrict access to specific buckets, you must add the new Amazon Linux bucket ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` to your policy (replace `$region` with the region where the endpoint is). For more information, see this topic in the AWS discussion forums. Announcement: Amazon Linux 2 now supports the ability to use HTTPS while connecting to package repositories .

Known issues

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that’s less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&‘()*+,-`. For more information, see UTF-8 encoding table and Unicode characters.

As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

6.3.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form `CommunityVersion-amzn-EmrVersion`. The `EmrVersion` starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.2.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.2.0	Conda env for emr notebook which includes jupyter enterprise gateway

Component	Version	Description
emr-s3-dist-cp	2.18.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.1.0	EMR S3Select Connector
emrfs	2.46.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.12.1	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.12.1	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	3.2.1-amzn-3.1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-3.1	HDFS node-level service for storing blocks.

Component	Version	Description
hadoop-hdfs-library	3.2.1-amzn-3.1	HDFS command-line client and library
hadoop-hdfs-namenode	3.2.1-amzn-3.1	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	3.2.1-amzn-3.1	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	3.2.1-amzn-3.1	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.2.1-amzn-3.1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.2.1-amzn-3.1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-3.1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.2.1-amzn-3.1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.2.1-amzn-3.1	Service for retrieving current and historical information for YARN applications.

Component	Version	Description
hbase-hmaster	2.2.6-amzn-1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.2.6-amzn-1	Service for serving one or more HBase regions.
hbase-client	2.2.6-amzn-1	HBase command-line client.
hbase-rest-server	2.2.6-amzn-1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	2.2.6-amzn-1	Service providing a Thrift endpoint to HBase.
hcatalog-client	3.1.2-amzn-4	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.2-amzn-4	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.2-amzn-4	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.2-amzn-4	Hive command line client.
hive-hbase	3.1.2-amzn-4	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-4	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.

Component	Version	Description
hive-server2	3.1.2-amzn-4	Service for accepting Hive queries as web requests.
hudi	0.7.0-amzn-0	Incremental processing framework to power data pipline at low latency and high efficiency.
hudi-presto	0.7.0-amzn-0	Bundle library for running Presto with Hudi.
hudi-prestosql	0.7.0-amzn-0	Bundle library for running PrestoSQL with Hudi.
hudi-spark	0.7.0-amzn-0	Bundle library for running Spark with Hudi.
hue-server	4.9.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.2.2	Multi-user server for Jupyter notebooks
livy-server	0.7.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.7.0	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68+	MariaDB database server.

Component	Version	Description
nvidia-cuda	10.1.243	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	The phoenix libraries for server and client
phoenix-query-server	5.0.0-HBase-2.0	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.245.1-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.245.1-amzn-0	Service for executing pieces of a query.
presto-client	0.245.1-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.

Component	Version	Description
prestosql-coordinator	350	Service for accepting queries and managing query execution among prestosql-workers.
prestosql-worker	350	Service for executing pieces of a query.
prestosql-client	350	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.1.1-amzn-0.1	Spark command-line clients.
spark-history-server	3.1.1-amzn-0.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	3.1.1-amzn-0.1	In-memory execution engine for YARN.
spark-yarn-slave	3.1.1-amzn-0.1	Apache Spark libraries needed by YARN slaves.
spark-rapids	0.4.1	Nvidia Spark RAPIDS plugin that accelerates Apache Spark with GPUs.

Component	Version	Description
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.4.1	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.41+	Apache HTTP server.
zeppelin-server	0.9.0	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

6.3.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see [Configure applications](#).

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see [Reconfigure an instance group in a running cluster](#).

emr-6.3.1 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN

Classifications	Description	Reconfiguration Actions
		services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-ses sion.properties settings for Kubernetes/Yarn session.	Restarts Flink history server.
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Restarts Flink history server.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts PhoenixQu eryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classifications	Description	Reconfiguration Actions
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	This classification should not be reconfigured.
hdfs-env	Change values in the HDFS environment.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat server.

Classifications	Description	Reconfiguration Actions
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat server.
hive	Amazon EMR-curated settings for Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.
hive-env	Change values in the Hive environment.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Not available.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.

Classifications	Description	Reconfiguration Actions
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Change values in the Hudi environment.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.

Classifications	Description	Reconfiguration Actions
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j	Change Livy log4j.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hb ase.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServer.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoDB)
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoDB)
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
prestosql-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoSQL)
prestosql-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoSQL)
prestosql-password-authenticator	Change values in Presto's password-authenticator.properties file.	Restarts Presto-Server (for PrestoSQL)
prestosql-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoSQL)
prestosql-node	Change values in PrestoSQL's node.properties file.	Not available.
prestosql-connector-blackhole	Change values in PrestoSQL's blackhole.properties file.	Not available.
prestosql-connector-cassandra	Change values in PrestoSQL's cassandra.properties file.	Not available.
prestosql-connector-hive	Change values in PrestoSQL's hive.properties file.	Restarts Presto-Server (for PrestoSQL)
prestosql-connector-jmx	Change values in PrestoSQL's jmx.properties file.	Not available.
prestosql-connector-kafka	Change values in PrestoSQL's kafka.properties file.	Not available.
prestosql-connector-localfile	Change values in PrestoSQL's localfile.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
prestosql-connector-memory	Change values in PrestoSQL's memory.properties file.	Not available.
prestosql-connector-mongodb	Change values in PrestoSQL's mongodb.properties file.	Not available.
prestosql-connector-mysql	Change values in PrestoSQL's mysql.properties file.	Not available.
prestosql-connector-postgresql	Change values in PrestoSQL's postgresql.properties file.	Not available.
prestosql-connector-raptor	Change values in PrestoSQL's raptor.properties file.	Not available.
prestosql-connector-redis	Change values in PrestoSQL's redis.properties file.	Not available.
prestosql-connector-redshift	Change values in PrestoSQL's redshift.properties file.	Not available.
prestosql-connector-tpch	Change values in PrestoSQL's tpch.properties file.	Not available.
prestosql-connector-tpcds	Change values in PrestoSQL's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.

Classifications	Description	Reconfiguration Actions
ranger-kms-log4j	Change values in kms-log4j .properties file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j	Change values in Spark's log4j.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restart Oozie and HiveServe r2.

Classifications	Description	Reconfiguration Actions
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts MapReduce -HistoryServer.
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Livy Server and MapReduce-HistoryS erver.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

Amazon EMR release 6.3.0

6.3.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, PrestoSQL, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.3.0	emr-6.2.1	emr-6.2.0	emr-6.1.1
AWS SDK for Java	1.11.977	1.11.880	1.11.880	1.11.828
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.12.10	2.12.10	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.12.1	1.11.2	1.11.2	1.11.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.6	2.2.6-amzn-0	2.2.6-amzn-0	2.2.5
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2

	emr-6.3.0	emr-6.2.1	emr-6.2.0	emr-6.1.1
Hudi	0.7.0-amzn-0	0.6.0-amzn-1	0.6.0-amzn-1	0.5.2-incubating-amzn-2
Hue	4.9.0	4.8.0	4.8.0	4.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	-
JupyterHub	1.2.2	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.7.0	1.6.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.0	5.2.0	5.2.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.245.1	0.238.3	0.238.3	0.232
Spark	3.1.1	3.0.1	3.0.1	3.0.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.3.1	2.3.1	2.1.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	350	343	343	338
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

6.3.0 release notes

The following release notes include information for Amazon EMR release 6.3.0. Changes are relative to 6.2.0.

Initial release date: May 12, 2021

Last updated date: August 9, 2021

Supported applications

- AWS SDK for Java version 1.11.977
- CloudWatch Sink version 2.1.0
- DynamoDB Connector version 4.16.0
- EMRFS version 2.46.0
- Amazon EMR Goodies version 3.2.0
- Amazon EMR Kinesis Connector version 3.5.0
- Amazon EMR Record Server version 2.0.0
- Amazon EMR Scripts version 2.5.0
- Flink version 1.12.1
- Ganglia version 3.7.2
- AWS Glue Hive Metastore Client version 3.2.0
- Hadoop version 3.2.1-amzn-3
- HBase version 2.2.6-amzn-1
- HBase-operator-tools 1.0.0
- HCatalog version 3.1.2-amzn-0
- Hive version 3.1.2-amzn-4
- Hudi version 0.7.0-amzn-0
- Hue version 4.9.0
- Java JDK version Corretto-8.282.08.1 (build 1.8.0_282-b08)
- JupyterHub version 1.2.0
- Livy version 0.7.0-incubating

- MXNet version 1.7.0
- Oozie version 5.2.1
- Phoenix version 5.0.0
- Pig version 0.17.0
- Presto version 0.245.1-amzn-0
- PrestoSQL version 350
- Apache Ranger KMS (multi-master transparent encryption) version 2.0.0
- ranger-plugins 2.0.1-amzn-0
- ranger-s3-plugin 1.1.0
- SageMaker Spark SDK version 1.4.1
- Scala version 2.12.10 (OpenJDK 64-Bit Server VM, Java 1.8.0_282)
- Spark version 3.1.1-amzn-0
- spark-rapids 0.4.1
- Sqoop version 1.4.7
- TensorFlow version 2.4.1
- tez version 0.9.2
- Zeppelin version 0.9.0
- Zookeeper version 3.4.14
- Connectors and drivers: DynamoDB Connector 4.16.0

New features

- Amazon EMR supports Amazon S3 Access Points, a feature of Amazon S3 that allows you to easily manage access for shared data lakes. Using your Amazon S3 Access Point alias, you can simplify your data access at scale on Amazon EMR. You can use Amazon S3 Access Points with all versions of Amazon EMR at no additional cost in all AWS regions where Amazon EMR is available. To learn more about Amazon S3 Access Points and Access Point aliases, see [Using a bucket-style alias for your access point]() in the *Amazon S3 User Guide*.
- New `DescribeReleaseLabel` and `ListReleaseLabel` API parameters provide Amazon EMR release label details. You can programmatically list releases available in the region where the API request is run, and list the available applications for a specific Amazon EMR release label. The

release label parameters also list Amazon EMR releases that support a specified application, such as Spark. This information can be used to programmatically launch Amazon EMR clusters. For example, you can launch a cluster using the latest release version from the `ListReleaseLabel` results. For more information, see DescribeReleaseLabel and ListReleaseLabels in the *Amazon EMR API Reference*.

- With Amazon EMR 6.3.0, you can launch a cluster that natively integrates with Apache Ranger. Apache Ranger is an open-source framework to enable, monitor, and manage comprehensive data security across the Hadoop platform. For more information, see Apache Ranger. With native integration, you can bring your own Apache Ranger to enforce fine-grained data access control on Amazon EMR. See Integrate Amazon EMR with Apache Ranger in the Amazon EMR Management Guide.
- Scoped managed policies: To align with AWS best practices, Amazon EMR has introduced v2 EMR-scoped default managed policies as replacements for policies that will be deprecated. See Amazon EMR Managed Policies.
- Instance Metadata Service (IMDS) V2 support status: For Amazon EMR 6.2 or later, Amazon EMR components use IMDSv2 for all IMDS calls. For IMDS calls in your application code, you can use both IMDSv1 and IMDSv2, or configure the IMDS to use only IMDSv2 for added security. If you disable IMDSv1 in earlier Amazon EMR 6.x releases, it causes cluster startup failure.

Changes, enhancements, and resolved issues

- This is a release to fix issues with Amazon EMR Scaling when it fails to scale up/scale down a cluster successfully or causes application failures.
- Fixed an issue where scaling requests failed for a large, highly utilized cluster when Amazon EMR on-cluster daemons were running health checking activities, such as gathering YARN node state and HDFS node state. This was happening because on-cluster daemons were not able to communicate the health status data of a node to internal Amazon EMR components.
- Improved EMR on-cluster daemons to correctly track the node states when IP addresses are reused to improve reliability during scaling operations.
- SPARK-29683. Fixed an issue where job failures occurred during cluster scale-down as Spark was assuming all available nodes were deny-listed.
- YARN-9011. Fixed an issue where job failures occurred due to a race condition in YARN decommissioning when cluster tried to scale up or down.
- Fixed issue with step or job failures during cluster scaling by ensuring that the node states are always consistent between the Amazon EMR on-cluster daemons and YARN/HDFS.

- Fixed an issue where cluster operations such as scale down and step submission failed for Amazon EMR clusters enabled with Kerberos authentication. This was because the Amazon EMR on-cluster daemon did not renew the Kerberos ticket, which is required to securely communicate with HDFS/YARN running on the primary node.
- Newer Amazon EMR releases fix the issue with a lower "Max open files" limit on older AL2 in Amazon EMR. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later now include a permanent fix with a higher "Max open files" setting.
- Spark SQL UI explain mode default changed from `extended` to `formatted` in [Spark 3.1](). Amazon EMR reverted it back to `extended` to include logical plan information in the Spark SQL UI. This can be reverted by setting `spark.sql.ui.explainMode` to `formatted`.
- The following commits were backported from the Spark master branch.

  - [SPARK-34752][BUILD] Bump Jetty to 9.4.37 to address CVE-2020-27223.

  - [SPARK-34534] Fix blockIds order when use FetchShuffleBlocks to fetch blocks.

  - [SPARK-34681] [SQL] Fix bug for full outer shuffled hash join when building left side with non-equal condition.

  - [SPARK-34497] [SQL] Fix built-in JDBC connection providers to restore JVM security context changes.
- To improve interoperability with Nvidia Spark RAPIDs plugin, Added workaround to address an issue preventing dynamic partition pruning from triggering when using Nvidia Spark RAPIDs with adaptive query execution disabled, see RAPIDS issue #1378 and RAPIDS issue ##1386. For details of the new configuration `spark.sql.optimizer.dynamicPartitionPruning.enforceBroadcastReuse`, see RAPIDS issue ##1386.
- The file output committer default algorithm has been changed from the v2 algorithm to the v1 algorithm in open source Spark 3.1. For more information, see this Amazon EMR optimizing Spark performance - dynamic partition pruning.
- Amazon EMR reverted to the v2 algorithm, the default used in prior Amazon EMR 6.x releases, to prevent performance regression. To restore the open source Spark 3.1 behavior, set `spark.hadoop.mapreduce.fileoutputcommitter.algorithm.version` to 1. Open source Spark made this change because task commit in file output committer algorithm v2 is not atomic, which can cause an output data correctness issue in some cases. However, task commit in algorithm v1 is also not atomic. In some scenarios task commit includes a delete performed before a rename. This can result in a silent data correctness issue.

- Fixed Managed Scaling issues in earlier Amazon EMR releases and made improvements so application failure rates are significantly reduced.
- Installed the AWS Java SDK Bundle on each new cluster. This is a single jar containing all service SDKs and their dependencies, instead of individual component jars. For more information, see Java SDK Bundled Dependency.

Known issues

- For Amazon EMR 6.3.0 and 6.2.0 private subnet clusters, you cannot access the Ganglia web UI. You will get an "access denied (403)" error. Other web UIs, such as Spark, Hue, JupyterHub, Zeppelin, Livy, and Tez are working normally. Ganglia web UI access on public subnet clusters are also working normally. To resolve this issue, restart httpd service on the primary node with `sudo systemctl restart httpd`. This issue is fixed in Amazon EMR 6.4.0.
- When AWS Glue Data Catalog is enabled, using Spark to access a AWS Glue DB with null string location URI may fail. This happens to earlier Amazon EMR releases, but SPARK-31709 (https://issues.apache.org/jira/browse/SPARK-31709) makes it apply to more cases. For example, when creating a table within the default AWS Glue DB whose location URI is a null string, `spark.sql("CREATE TABLE mytest (key string) location '/table_path';")` fails with the message, "Cannot create a Path from an empty string." To work around this, manually set a location URI of your AWS Glue databases, then create tables within these databases using Spark.
- In Amazon EMR 6.3.0, PrestoSQL has upgraded from version 343 to version 350. There are two security related changes from the open source that relate to this version change. File-based catalog access control is changed from `deny` to `allow` when table, schema, or session property rules are not defined. Also, file-based system access control is changed to support files without catalog rules defined. In this case, all access to catalogs is allowed.

  For more information, see Release 344 (9 Oct 2020).
- Note that the Hadoop user directory (/home/hadoop) is readable by everyone. It has Unix 755 (drwxr-xr-x) directory permissions to allow read access by frameworks like Hive. You can put files in /home/hadoop and its subdirectories, but be aware of the permissions on those directories to protect sensitive information.
- **Lower "Max open files" limit on older AL2 [fixed in newer releases].** Amazon EMR releases: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0, and emr-6.2.0 are based on older versions ofAmazon Linux 2 (AL2), which have a lower ulimit setting for "Max open files" when Amazon EMR clusters are created with the default AMI. Amazon EMR releases 5.30.1, 5.30.2,

5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later include a permanent fix with a higher "Max open files" setting. Releases with the lower open file limit causes a "Too many open files" error when submitting Spark job. In the impacted releases, the Amazon EMR default AMI has a default ulimit setting of 4096 for "Max open files," which is lower than the 65536 file limit in the latestAmazon Linux 2 AMI. The lower ulimit setting for "Max open files" causes Spark job failure when the Spark driver and executor try to open more than 4096 files. To fix the issue, Amazon EMR has a bootstrap action (BA) script that adjusts the ulimit setting at cluster creation.

If you are using an older Amazon EMR version that doesn't have the permanent fix for this issue, the following workaround lets you to explicitly set the instance-controller ulimit to a maximum of 65536 files.

Explicitly set a ulimit from the command line

1. Edit `/etc/systemd/system/instance-controller.service` to add the following parameters to Service section.

   `LimitNOFILE=65536`

   `LimitNPROC=65536`
2. Restart InstanceController

   `$ sudo systemctl daemon-reload`

   `$ sudo systemctl restart instance-controller`

Set a ulimit using bootstrap action (BA)

You can also use a bootstrap action (BA) script to configure the instance-controller ulimit to 65536 files at cluster creation.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
```

```
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- **Important**

  EMR clusters that run Amazon Linux or Amazon Linux 2 Amazon Machine Images (AMIs) use default Amazon Linux behavior, and do not automatically download and install important and critical kernel updates that require a reboot. This is the same behavior as other Amazon EC2 instances that run the default Amazon Linux AMI. If new Amazon Linux software updates that require a reboot (such as kernel, NVIDIA, and CUDA updates) become available after an Amazon EMR release becomes available, EMR cluster instances that run the default AMI do not automatically download and install those updates. To get kernel updates, you can customize your Amazon EMR AMI to use the latest Amazon Linux AMI.

- To use Spark actions with Apache Oozie, you must add the following configuration to your Oozie `workflow.xml` file. Otherwise, several critical libraries such as Hadoop and EMRFS will be missing from the classpath of the Spark executors that Oozie launches.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that

there are 14 other non-control characters: `!"#$%&‘()*+,-`. For more information, see UTF-8 encoding table and Unicode characters.

As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

6.3.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.2.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.

Component	Version	Description
emr-notebook-env	1.2.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.18.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.1.0	EMR S3Select Connector
emrfs	2.46.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.12.1	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.12.1	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	3.2.1-amzn-3	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.

Component	Version	Description
hadoop-hdfs-datanode	3.2.1-amzn-3	HDFS node-level service for storing blocks.
hadoop-hdfs-library	3.2.1-amzn-3	HDFS command-line client and library
hadoop-hdfs-namenode	3.2.1-amzn-3	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	3.2.1-amzn-3	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	3.2.1-amzn-3	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.2.1-amzn-3	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.2.1-amzn-3	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-3	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.2.1-amzn-3	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.2.1-amzn-3	Service for retrieving current and historical information for YARN applications.

Component	Version	Description
hbase-hmaster	2.2.6-amzn-1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.2.6-amzn-1	Service for serving one or more HBase regions.
hbase-client	2.2.6-amzn-1	HBase command-line client.
hbase-rest-server	2.2.6-amzn-1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	2.2.6-amzn-1	Service providing a Thrift endpoint to HBase.
hcatalog-client	3.1.2-amzn-4	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.2-amzn-4	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.2-amzn-4	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.2-amzn-4	Hive command line client.
hive-hbase	3.1.2-amzn-4	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-4	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.

Component	Version	Description
hive-server2	3.1.2-amzn-4	Service for accepting Hive queries as web requests.
hudi	0.7.0-amzn-0	Incremental processing framework to power data pipline at low latency and high efficiency.
hudi-presto	0.7.0-amzn-0	Bundle library for running Presto with Hudi.
hudi-prestosql	0.7.0-amzn-0	Bundle library for running PrestoSQL with Hudi.
hudi-spark	0.7.0-amzn-0	Bundle library for running Spark with Hudi.
hue-server	4.9.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.2.2	Multi-user server for Jupyter notebooks
livy-server	0.7.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.7.0	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68+	MariaDB database server.

Component	Version	Description
nvidia-cuda	10.1.243	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	The phoenix libraries for server and client
phoenix-query-server	5.0.0-HBase-2.0	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.245.1-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.245.1-amzn-0	Service for executing pieces of a query.
presto-client	0.245.1-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.

Component	Version	Description
prestosql-coordinator	350	Service for accepting queries and managing query execution among prestosql-workers.
prestosql-worker	350	Service for executing pieces of a query.
prestosql-client	350	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.1.1-amzn-0	Spark command-line clients.
spark-history-server	3.1.1-amzn-0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	3.1.1-amzn-0	In-memory execution engine for YARN.
spark-yarn-slave	3.1.1-amzn-0	Apache Spark libraries needed by YARN slaves.
spark-rapids	0.4.1	Nvidia Spark RAPIDS plugin that accelerates Apache Spark with GPUs.

Component	Version	Description
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.4.1	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.41+	Apache HTTP server.
zeppelin-server	0.9.0	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

6.3.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-6.3.0 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN

Classifications	Description	Reconfiguration Actions
		services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-ses sion.properties settings for Kubernetes/Yarn session.	Restarts Flink history server.
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Restarts Flink history server.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts PhoenixQu eryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classifications	Description	Reconfiguration Actions
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	This classification should not be reconfigured.
hdfs-env	Change values in the HDFS environment.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat server.

Classifications	Description	Reconfiguration Actions
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat server.
hive	Amazon EMR-curated settings for Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.
hive-env	Change values in the Hive environment.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Not available.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.

Classifications	Description	Reconfiguration Actions
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Change values in the Hudi environment.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.

Classifications	Description	Reconfiguration Actions
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j	Change Livy log4j.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hb ase.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServer.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoDB)
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoDB)
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
prestosql-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoSQL)
prestosql-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoSQL)
prestosql-password-authenticator	Change values in Presto's password-authenticator.properties file.	Restarts Presto-Server (for PrestoSQL)
prestosql-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoSQL)
prestosql-node	Change values in PrestoSQL's node.properties file.	Not available.
prestosql-connector-blackhole	Change values in PrestoSQL's blackhole.properties file.	Not available.
prestosql-connector-cassandra	Change values in PrestoSQL's cassandra.properties file.	Not available.
prestosql-connector-hive	Change values in PrestoSQL's hive.properties file.	Restarts Presto-Server (for PrestoSQL)
prestosql-connector-jmx	Change values in PrestoSQL's jmx.properties file.	Not available.
prestosql-connector-kafka	Change values in PrestoSQL's kafka.properties file.	Not available.
prestosql-connector-localfile	Change values in PrestoSQL's localfile.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
prestosql-connector-memory	Change values in PrestoSQL's memory.properties file.	Not available.
prestosql-connector-mongodb	Change values in PrestoSQL's mongodb.properties file.	Not available.
prestosql-connector-mysql	Change values in PrestoSQL's mysql.properties file.	Not available.
prestosql-connector-postgresql	Change values in PrestoSQL's postgresql.properties file.	Not available.
prestosql-connector-raptor	Change values in PrestoSQL's raptor.properties file.	Not available.
prestosql-connector-redis	Change values in PrestoSQL's redis.properties file.	Not available.
prestosql-connector-redshift	Change values in PrestoSQL's redshift.properties file.	Not available.
prestosql-connector-tpch	Change values in PrestoSQL's tpch.properties file.	Not available.
prestosql-connector-tpcds	Change values in PrestoSQL's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.

Classifications	Description	Reconfiguration Actions
ranger-kms-log4j	Change values in kms-log4j .properties file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j	Change values in Spark's log4j.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restart Oozie and HiveServe r2.

Classifications	Description	Reconfiguration Actions
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts MapReduce -HistoryServer.
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Livy Server and MapReduce-HistoryS erver.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

Amazon EMR release 6.2.1

6.2.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, PrestoSQL, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.2.1	emr-6.2.0	emr-6.1.1	emr-6.1.0
AWS SDK for Java	1.11.880	1.11.880	1.11.828	1.11.828
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.12.10	2.12.10	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.2	1.11.2	1.11.0	1.11.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.6-amzn-0	2.2.6-amzn-0	2.2.5	2.2.5
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2

	emr-6.2.1	emr-6.2.0	emr-6.1.1	emr-6.1.0
Hudi	0.6.0-amzn-1	0.6.0-amzn-1	0.5.2-inc ubating-amzn-2	0.5.2-inc ubating-amzn-2
Hue	4.8.0	4.8.0	4.7.1	4.7.1
Iceberg	-	-	-	-
JupyterEn terpriseGateway	2.1.0	2.1.0	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.6.0	1.6.0
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.238.3	0.238.3	0.232	0.232
Spark	3.0.1	3.0.1	3.0.0	3.0.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.3.1	2.3.1	2.1.0	2.1.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQ L)	343	343	338	338
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

6.2.1 release notes

This is a release to fix issues with Amazon EMR Scaling when it fails to scale up/scale down a cluster successfully or causes application failures.

Changes, Enhancements, and Resolved Issues

- Fixed an issue where scaling requests failed for a large, highly utilized cluster when Amazon EMR on-cluster daemons were running health checking activities, such as gathering YARN node state and HDFS node state. This was happening because on-cluster daemons were not able to communicate the health status data of a node to internal Amazon EMR components.
- Improved EMR on-cluster daemons to correctly track the node states when IP addresses are reused to improve reliability during scaling operations.
- SPARK-29683. Fixed an issue where job failures occurred during cluster scale-down as Spark was assuming all available nodes were deny-listed.
- YARN-9011. Fixed an issue where job failures occurred due to a race condition in YARN decommissioning when cluster tried to scale up or down.
- Fixed issue with step or job failures during cluster scaling by ensuring that the node states are always consistent between the Amazon EMR on-cluster daemons and YARN/HDFS.
- Fixed an issue where cluster operations such as scale down and step submission failed for Amazon EMR clusters enabled with Kerberos authentication. This was because the Amazon EMR on-cluster daemon did not renew the Kerberos ticket, which is required to securely communicate with HDFS/YARN running on the primary node.
- Newer Amazon EMR releases fix the issue with a lower "Max open files" limit on older AL2 in Amazon EMR. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later now include a permanent fix with a higher "Max open files" setting.
- HTTPS is now enabled by default for Amazon Linux repositories. If you are using an Amazon S3 VPCE policy to restrict access to specific buckets, you must add the new Amazon Linux bucket ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` to your policy (replace `$region` with the region where the endpoint is). For more information, see this topic in the AWS discussion forums. Announcement: Amazon Linux 2 now supports the ability to use HTTPS while connecting to package repositories .

Known issues

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&'()*+,-`. For more information, see [UTF-8 encoding table and Unicode characters](https://www.utf8-chartable.de/).

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

6.2.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK

Component	Version	Description
emr-ddb	4.16.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.1.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.0.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.16.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.0.0	EMR S3Select Connector
emrfs	2.44.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.11.2	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.11.2	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.

Component	Version	Description
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	3.2.1-amzn-2.1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-2.1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	3.2.1-amzn-2.1	HDFS command-line client and library
hadoop-hdfs-namenode	3.2.1-amzn-2.1	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	3.2.1-amzn-2.1	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	3.2.1-amzn-2.1	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.2.1-amzn-2.1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.2.1-amzn-2.1	MapReduce execution engine libraries for running a MapReduce application.

Component	Version	Description
hadoop-yarn-nodemanager	3.2.1-amzn-2.1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.2.1-amzn-2.1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.2.1-amzn-2.1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	2.2.6-amzn-0	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.2.6-amzn-0	Service for serving one or more HBase regions.
hbase-client	2.2.6-amzn-0	HBase command-line client.
hbase-rest-server	2.2.6-amzn-0	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	2.2.6-amzn-0	Service providing a Thrift endpoint to HBase.
hcatalog-client	3.1.2-amzn-3	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.2-amzn-3	Service providing HCatalog, a table and storage management layer for distributed applications.

Component	Version	Description
hcatalog-webhcat-server	3.1.2-amzn-3	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.2-amzn-3	Hive command line client.
hive-hbase	3.1.2-amzn-3	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-3	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	3.1.2-amzn-3	Service for accepting Hive queries as web requests.
hudi	0.6.0-amzn-1	Incremental processing framework to power data pipline at low latency and high efficiency.
hudi-presto	0.6.0-amzn-1	Bundle library for running Presto with Hudi.
hudi-prestosql	0.6.0-amzn-1	Bundle library for running PrestoSQL with Hudi.
hudi-spark	0.6.0-amzn-1	Bundle library for running Spark with Hudi.
hue-server	4.8.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.1.0	Multi-user server for Jupyter notebooks

Component	Version	Description
livy-server	0.7.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.7.0	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.64+	MariaDB database server.
nvidia-cuda	10.1.243	Nvidia drivers and Cuda toolkit
oozie-client	5.2.0	Oozie command-line client.
oozie-server	5.2.0	Service for accepting Oozie workflow requests.
opencv	4.4.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	The phoenix libraries for server and client
phoenix-query-server	5.0.0-HBase-2.0	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.238.3-amzn-1	Service for accepting queries and managing query execution among presto-wo rkers.

Component	Version	Description
presto-worker	0.238.3-amzn-1	Service for executing pieces of a query.
presto-client	0.238.3-amzn-1	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
prestosql-coordinator	343	Service for accepting queries and managing query execution among prestosql-workers.
prestosql-worker	343	Service for executing pieces of a query.
prestosql-client	343	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
pig-client	0.17.0	Pig command-line client.
r	3.4.3	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.0.1-amzn-0.1	Spark command-line clients.
spark-history-server	3.0.1-amzn-0.1	Web UI for viewing logged events for the lifetime of a completed Spark application.

Component	Version	Description
spark-on-yarn	3.0.1-amzn-0.1	In-memory execution engine for YARN.
spark-yarn-slave	3.0.1-amzn-0.1	Apache Spark libraries needed by YARN slaves.
spark-rapids	0.2.0	Nvidia Spark RAPIDS plugin that accelerates Apache Spark with GPUs.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.3.1	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.41+	Apache HTTP server.
zeppelin-server	0.9.0-preview1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

6.2.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-6.2.1 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceManager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Not available.
flink-log4j	Change Flink log4j.properties settings.	Not available.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.	Not available.
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Not available.

Classifications	Description	Reconfiguration Actions
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services SecondaryNamenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.

Classifications	Description	Reconfiguration Actions
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	This classification should not be reconfigured.
hdfs-env	Change values in the HDFS environment.	Restarts Hadoop HDFS ZKFC.

Classifications	Description	Reconfiguration Actions
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat server.
hive	Amazon EMR-curated settings for Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
hive-env	Change values in the Hive environment.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Not available.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.

Classifications	Description	Reconfiguration Actions
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Change values in the Hudi environment.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j	Change Livy log4j.properties settings.	Restarts Livy Server.

Classifications	Description	Reconfiguration Actions
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServer.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.

Classifications	Description	Reconfiguration Actions
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoDB)
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoDB)
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
prestosql-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoSQL)
prestosql-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoSQL)
prestosql-password-authenticator	Change values in Presto's password-authenticator.properties file.	Restarts Presto-Server (for PrestoSQL)

Classifications	Description	Reconfiguration Actions
prestosql-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoSQL)
prestosql-node	Change values in PrestoSQL's node.properties file.	Not available.
prestosql-connector-blackhole	Change values in PrestoSQL's blackhole.properties file.	Not available.
prestosql-connector-cassandra	Change values in PrestoSQL's cassandra.properties file.	Not available.
prestosql-connector-hive	Change values in PrestoSQL's hive.properties file.	Restarts Presto-Server (for PrestoSQL)
prestosql-connector-jmx	Change values in PrestoSQL's jmx.properties file.	Not available.
prestosql-connector-kafka	Change values in PrestoSQL's kafka.properties file.	Not available.
prestosql-connector-localfile	Change values in PrestoSQL's localfile.properties file.	Not available.
prestosql-connector-memory	Change values in PrestoSQL's memory.properties file.	Not available.
prestosql-connector-mongodb	Change values in PrestoSQL's mongodb.properties file.	Not available.
prestosql-connector-mysql	Change values in PrestoSQL's mysql.properties file.	Not available.
prestosql-connector-postgresql	Change values in PrestoSQL's postgresql.properties file.	Not available.
prestosql-connector-raptor	Change values in PrestoSQL's raptor.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
prestosql-connector-redis	Change values in PrestoSQL's redis.properties file.	Not available.
prestosql-connector-redshift	Change values in PrestoSQL's redshift.properties file.	Not available.
prestosql-connector-tpch	Change values in PrestoSQL's tpch.properties file.	Not available.
prestosql-connector-tpcds	Change values in PrestoSQL's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-log4j	Change values in kms-log4j .properties file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.

Classifications	Description	Reconfiguration Actions
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j	Change values in Spark's log4j.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restart Oozie.
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts MapReduce -HistoryServer.
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Livy Server and MapReduce-HistoryS erver.

Classifications	Description	Reconfiguration Actions
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

Amazon EMR release 6.2.0

6.2.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, PrestoSQL, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.2.0	emr-6.1.1	emr-6.1.0	emr-6.0.1
AWS SDK for Java	1.11.880	1.11.828	1.11.828	1.11.711
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7

	emr-6.2.0	emr-6.1.1	emr-6.1.0	emr-6.0.1
Scala	2.12.10	2.12.10	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.2	1.11.0	1.11.0	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.6-amzn-0	2.2.5	2.2.5	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.6.0-amzn-1	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1
Hue	4.8.0	4.7.1	4.7.1	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.0.0
Livy	0.7.0	0.7.0	0.7.0	0.6.0
MXNet	1.7.0	1.6.0	1.6.0	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.2.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0

	emr-6.2.0	emr-6.1.1	emr-6.1.0	emr-6.0.1
Pig	0.17.0	0.17.0	0.17.0	-
Presto	0.238.3	0.232	0.232	0.230
Spark	3.0.1	3.0.0	3.0.0	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	-
TensorFlow	2.3.1	2.1.0	2.1.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	343	338	338	-
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

6.2.0 release notes

The following release notes include information for Amazon EMR release 6.2.0. Changes are relative to 6.1.0.

Initial release date: Dec 09, 2020

Last updated date: Oct 04, 2021

Supported applications

- AWS SDK for Java version 1.11.828
- emr-record-server version 1.7.0
- Flink version 1.11.2
- Ganglia version 3.7.2
- Hadoop version 3.2.1-amzn-1
- HBase version 2.2.6-amzn-0
- HBase-operator-tools 1.0.0

- HCatalog version 3.1.2-amzn-0
- Hive version 3.1.2-amzn-3
- Hudi version 0.6.0-amzn-1
- Hue version 4.8.0
- JupyterHub version 1.1.0
- Livy version 0.7.0
- MXNet version 1.7.0
- Oozie version 5.2.0
- Phoenix version 5.0.0
- Pig version 0.17.0
- Presto version 0.238.3-amzn-1
- PrestoSQL version 343
- Spark version 3.0.1-amzn-0
- spark-rapids 0.2.0
- TensorFlow version 2.3.1
- Zeppelin version 0.9.0-preview1
- Zookeeper version 3.4.14
- Connectors and drivers: DynamoDB Connector 4.16.0

New features

- HBase: Removed rename in commit phase and added persistent HFile tracking. See Persistent HFile Tracking in the *Amazon EMR Release Guide*.
- HBase: Backported Create a config that forces to cache blocks on compaction.
- PrestoDB: Improvements to Dynamic Partition Pruning. Rule-based Join Reorder works on non-partitioned data.
- Scoped managed policies: To align with AWS best practices, Amazon EMR has introduced v2 EMR-scoped default managed policies as replacements for policies that will be deprecated. See Amazon EMR Managed Policies.
- Instance Metadata Service (IMDS) V2 support status: For Amazon EMR 6.2 or later, Amazon EMR components use IMDSv2 for all IMDS calls. For IMDS calls in your application code, you can use

both IMDSv1 and IMDSv2, or configure the IMDS to use only IMDSv2 for added security. If you disable IMDSv1 in earlier Amazon EMR 6.x releases, it causes cluster startup failure.

Changes, enhancements, and resolved issues

- This is a release to fix issues with Amazon EMR Scaling when it fails to scale up/scale down a cluster successfully or causes application failures.
- Fixed an issue where scaling requests failed for a large, highly utilized cluster when Amazon EMR on-cluster daemons were running health checking activities, such as gathering YARN node state and HDFS node state. This was happening because on-cluster daemons were not able to communicate the health status data of a node to internal Amazon EMR components.
- Improved EMR on-cluster daemons to correctly track the node states when IP addresses are reused to improve reliability during scaling operations.
- SPARK-29683. Fixed an issue where job failures occurred during cluster scale-down as Spark was assuming all available nodes were deny-listed.
- YARN-9011. Fixed an issue where job failures occurred due to a race condition in YARN decommissioning when cluster tried to scale up or down.
- Fixed issue with step or job failures during cluster scaling by ensuring that the node states are always consistent between the Amazon EMR on-cluster daemons and YARN/HDFS.
- Fixed an issue where cluster operations such as scale down and step submission failed for Amazon EMR clusters enabled with Kerberos authentication. This was because the Amazon EMR on-cluster daemon did not renew the Kerberos ticket, which is required to securely communicate with HDFS/YARN running on the primary node.
- Newer Amazon EMR releases fix the issue with a lower "Max open files" limit on older AL2 in Amazon EMR. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later now include a permanent fix with a higher "Max open files" setting.
- Spark: Performance improvements in Spark runtime.

Known issues

- Amazon EMR 6.2 has incorrect permissions set on the /etc/cron.d/libinstance-controller-java file in EMR 6.2.0. Permissions on the file are 645 (-rw-r--r-x), when they should be 644 (-rw-r--r--). As a result, Amazon EMR version 6.2 does not log instance-state logs, and the /emr/instance-logs directory is empty. This issue is fixed in Amazon EMR 6.3.0 and later.

To work around this issue, run the following script as a bootstrap action at cluster launch.

```
#!/bin/bash
sudo chmod 644 /etc/cron.d/libinstance-controller-java
```

- For Amazon EMR 6.2.0 and 6.3.0 private subnet clusters, you cannot access the Ganglia web UI. You will get an "access denied (403)" error. Other web UIs, such as Spark, Hue, JupyterHub, Zeppelin, Livy, and Tez are working normally. Ganglia web UI access on public subnet clusters are also working normally. To resolve this issue, restart httpd service on the primary node with `sudo systemctl restart httpd`. This issue is fixed in Amazon EMR 6.4.0.
- There is an issue in Amazon EMR 6.2.0 where httpd continuously fails, causing Ganglia to be unavailable. You get a "cannot connect to the server" error. To fix a cluster that is already running with this issue, SSH to the cluster primary node and add the line `Listen 80` to the file `httpd.conf` located at `/etc/httpd/conf/httpd.conf`. This issue is fixed in Amazon EMR 6.3.0.
- HTTPD fails on EMR 6.2.0 clusters when you use a security configuration. This makes the Ganglia web application user interface unavailable. To access the Ganglia web application user interface, add `Listen 80` to the `/etc/httpd/conf/httpd.conf` file on the primary node of your cluster. For information about connecting to your cluster, see Connect to the Primary Node Using SSH.

  EMR Notebooks also fail to establish a connection with EMR 6.2.0 clusters when you use a security configuration. The notebook will fail to list kernels and submit Spark jobs. We recommend that you use EMR Notebooks with another version of Amazon EMR instead.
- **Lower "Max open files" limit on older AL2 [fixed in newer releases].** Amazon EMR releases: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0, and emr-6.2.0 are based on older versions ofAmazon Linux 2 (AL2), which have a lower ulimit setting for "Max open files" when Amazon EMR clusters are created with the default AMI. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later include a permanent fix with a higher "Max open files" setting. Releases with the lower open file limit causes a "Too many open files" error when submitting Spark job. In the impacted releases, the Amazon EMR default AMI has a default ulimit setting of 4096 for "Max open files," which is lower than the 65536 file limit in the latestAmazon Linux 2 AMI. The lower ulimit setting for "Max open files" causes Spark job failure when the Spark driver and executor try to open more than 4096 files. To fix the issue, Amazon EMR has a bootstrap action (BA) script that adjusts the ulimit setting at cluster creation.

If you are using an older Amazon EMR version that doesn't have the permanent fix for this issue, the following workaround lets you to explicitly set the instance-controller ulimit to a maximum of 65536 files.

Explicitly set a ulimit from the command line

1. Edit `/etc/systemd/system/instance-controller.service` to add the following parameters to Service section.

   `LimitNOFILE=65536`

   `LimitNPROC=65536`
2. Restart InstanceController

   `$ sudo systemctl daemon-reload`

   `$ sudo systemctl restart instance-controller`

Set a ulimit using bootstrap action (BA)

You can also use a bootstrap action (BA) script to configure the instance-controller ulimit to 65536 files at cluster creation.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
```

```
sudo systemctl daemon-reload
```

- **Important**

  Amazon EMR 6.1.0 and 6.2.0 include a performance issue that can critically affect all Hudi insert, upsert, and delete operations. If you plan to use Hudi with Amazon EMR 6.1.0 or 6.2.0, you should contact AWS support to obtain a patched Hudi RPM.

- **Important**

  EMR clusters that run Amazon Linux or Amazon Linux 2 Amazon Machine Images (AMIs) use default Amazon Linux behavior, and do not automatically download and install important and critical kernel updates that require a reboot. This is the same behavior as other Amazon EC2 instances that run the default Amazon Linux AMI. If new Amazon Linux software updates that require a reboot (such as kernel, NVIDIA, and CUDA updates) become available after an Amazon EMR release becomes available, EMR cluster instances that run the default AMI do not automatically download and install those updates. To get kernel updates, you can customize your Amazon EMR AMI to use the latest Amazon Linux AMI.

- Amazon EMR 6.2.0 Maven artifacts are not published. They will be published with a future release of Amazon EMR.
- Persistent HFile tracking using the HBase storefile system table does not support the HBase region replication feature. For more information about HBase region replication, see Timeline-consistent High Available Reads.
- Amazon EMR 6.x and EMR 5.x Hive bucketing version differences

  EMR 5.x uses OOS Apache Hive 2, while in EMR 6.x uses OOS Apache Hive 3. The open source Hive2 uses Bucketing version 1, while open source Hive3 uses Bucketing version 2. This bucketing version difference between Hive 2 (EMR 5.x) and Hive 3 (EMR 6.x) means Hive bucketing hashing functions differently. See the example below.

  The following table is an example created in EMR 6.x and EMR 5.x, respectively.

```
-- Using following LOCATION in EMR 6.x
CREATE TABLE test_bucketing (id INT, desc STRING)
PARTITIONED BY (day STRING)
CLUSTERED BY(id) INTO 128 BUCKETS
```

```
LOCATION 's3://your-own-s3-bucket/emr-6-bucketing/';

-- Using following LOCATION in EMR 5.x
LOCATION 's3://your-own-s3-bucket/emr-5-bucketing/';
```

Inserting the same data in both EMR 6.x and EMR 5.x.

```
INSERT INTO test_bucketing PARTITION (day='01') VALUES(66, 'some_data');
INSERT INTO test_bucketing PARTITION (day='01') VALUES(200, 'some_data');
```

Checking the S3 location, shows the bucketing file name is different, because the hashing function is different between EMR 6.x (Hive 3) and EMR 5.x (Hive 2).

```
[hadoop@ip-10-0-0-122 ~]$ aws s3 ls s3://your-own-s3-bucket/emr-6-bucketing/day=01/
2020-10-21 20:35:16         13 000025_0
2020-10-21 20:35:22         14 000121_0
[hadoop@ip-10-0-0-122 ~]$ aws s3 ls s3://your-own-s3-bucket/emr-5-bucketing/day=01/
2020-10-21 20:32:07         13 000066_0
2020-10-21 20:32:51         14 000072_0
```

You can also see the version difference by running the following command in Hive CLI in EMR 6.x. Note that it returns bucketing version 2.

```
hive> DESCRIBE FORMATTED test_bucketing;
...
Table Parameters:
    bucketing_version       2
...
```

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as `hadoop` user to the lead primary node of the EMR cluster with multiple primary nodes.

- Run the following command to renew Kerberos ticket for hadoop user.

```
kinit -kt <keytab_file> <principal>
```

Typically, the keytab file is located at /etc/hadoop.keytab and the principal is in the form of hadoop/<hostname>@<REALM>.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, s3://bucket/table/p=a is a prefix of s3://bucket/table/p=a b.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in s3://bucket/table/p=a b falls into this category. Note that there are 14 other non-control characters: !"#$%&‘()*+,-. For more information, see UTF-8 encoding table and Unicode characters.

  As a workaround to this issue, set the spark.sql.sources.fastS3PartitionDiscovery.enabled configuration to false in the spark-defaults classification.

6.2.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with emr or aws. Big-data application packages in the

most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form `CommunityVersion-amzn-EmrVersion`. The `EmrVersion` starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.1.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.0.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.16.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.0.0	EMR S3Select Connector
emrfs	2.44.0	Amazon S3 connector for Hadoop ecosystem applicati ons.

Component	Version	Description
flink-client	1.11.2	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.11.2	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	3.2.1-amzn-2	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-2	HDFS node-level service for storing blocks.
hadoop-hdfs-library	3.2.1-amzn-2	HDFS command-line client and library
hadoop-hdfs-namenode	3.2.1-amzn-2	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	3.2.1-amzn-2	HDFS service for managing the Hadoop filesystem journal on HA clusters.

Component	Version	Description
hadoop-httpfs-server	3.2.1-amzn-2	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.2.1-amzn-2	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.2.1-amzn-2	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-2	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.2.1-amzn-2	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.2.1-amzn-2	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	2.2.6-amzn-0	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.2.6-amzn-0	Service for serving one or more HBase regions.
hbase-client	2.2.6-amzn-0	HBase command-line client.
hbase-rest-server	2.2.6-amzn-0	Service providing a RESTful HTTP endpoint for HBase.

Component	Version	Description
hbase-thrift-server	2.2.6-amzn-0	Service providing a Thrift endpoint to HBase.
hcatalog-client	3.1.2-amzn-3	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.2-amzn-3	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.2-amzn-3	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.2-amzn-3	Hive command line client.
hive-hbase	3.1.2-amzn-3	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-3	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	3.1.2-amzn-3	Service for accepting Hive queries as web requests.
hudi	0.6.0-amzn-1	Incremental processing framework to power data pipline at low latency and high efficiency.
hudi-presto	0.6.0-amzn-1	Bundle library for running Presto with Hudi.

Component	Version	Description
hudi-prestosql	0.6.0-amzn-1	Bundle library for running PrestoSQL with Hudi.
hudi-spark	0.6.0-amzn-1	Bundle library for running Spark with Hudi.
hue-server	4.8.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.1.0	Multi-user server for Jupyter notebooks
livy-server	0.7.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.7.0	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.64+	MariaDB database server.
nvidia-cuda	10.1.243	Nvidia drivers and Cuda toolkit
oozie-client	5.2.0	Oozie command-line client.
oozie-server	5.2.0	Service for accepting Oozie workflow requests.
opencv	4.4.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	The phoenix libraries for server and client

Component	Version	Description
phoenix-query-server	5.0.0-HBase-2.0	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.238.3-amzn-1	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.238.3-amzn-1	Service for executing pieces of a query.
presto-client	0.238.3-amzn-1	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
prestosql-coordinator	343	Service for accepting queries and managing query execution among prestosql-workers.
prestosql-worker	343	Service for executing pieces of a query.
prestosql-client	343	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
pig-client	0.17.0	Pig command-line client.

Component	Version	Description
r	3.4.3	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.0.1-amzn-0	Spark command-line clients.
spark-history-server	3.0.1-amzn-0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	3.0.1-amzn-0	In-memory execution engine for YARN.
spark-yarn-slave	3.0.1-amzn-0	Apache Spark libraries needed by YARN slaves.
spark-rapids	0.2.0	Nvidia Spark RAPIDS plugin that accelerates Apache Spark with GPUs.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.3.1	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.41+	Apache HTTP server.

Component	Version	Description
zeppelin-server	0.9.0-preview1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

6.2.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-6.2.0 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.

Classifications	Description	Reconfiguration Actions
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.

Classifications	Description	Reconfiguration Actions
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Not available.
flink-log4j	Change Flink log4j.properties settings.	Not available.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.	Not available.
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Not available.

Classifications	Description	Reconfiguration Actions
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services SecondaryNamenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.

Classifications	Description	Reconfiguration Actions
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	This classification should not be reconfigured.
hdfs-env	Change values in the HDFS environment.	Restarts Hadoop HDFS ZKFC.

Classifications	Description	Reconfiguration Actions
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat server.
hive	Amazon EMR-curated settings for Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
hive-env	Change values in the Hive environment.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Not available.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.

Classifications	Description	Reconfiguration Actions
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Change values in the Hudi environment.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j	Change Livy log4j.properties settings.	Restarts Livy Server.

Classifications	Description	Reconfiguration Actions
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServer.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.

Classifications	Description	Reconfiguration Actions
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoDB)
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoDB)
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
prestosql-log	Change values in Presto's log.properties file.	Restarts Presto-Server (for PrestoSQL)
prestosql-config	Change values in Presto's config.properties file.	Restarts Presto-Server (for PrestoSQL)
prestosql-password-authenticator	Change values in Presto's password-authenticator.properties file.	Restarts Presto-Server (for PrestoSQL)

Classifications	Description	Reconfiguration Actions
prestosql-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server (for PrestoSQL)
prestosql-node	Change values in PrestoSQL's node.properties file.	Not available.
prestosql-connector-blackho le	Change values in PrestoSQL's blackhole.properties file.	Not available.
prestosql-connector-cassand ra	Change values in PrestoSQL's cassandra.properties file.	Not available.
prestosql-connector-hive	Change values in PrestoSQL's hive.properties file.	Restarts Presto-Server (for PrestoSQL)
prestosql-connector-jmx	Change values in PrestoSQL's jmx.properties file.	Not available.
prestosql-connector-kafka	Change values in PrestoSQL's kafka.properties file.	Not available.
prestosql-connector-localfile	Change values in PrestoSQL's localfile.properties file.	Not available.
prestosql-connector-memory	Change values in PrestoSQL's memory.properties file.	Not available.
prestosql-connector-mongodb	Change values in PrestoSQL's mongodb.properties file.	Not available.
prestosql-connector-mysql	Change values in PrestoSQL's mysql.properties file.	Not available.
prestosql-connector-postgre sql	Change values in PrestoSQL's postgresql.properties file.	Not available.
prestosql-connector-raptor	Change values in PrestoSQL's raptor.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
prestosql-connector-redis	Change values in PrestoSQL's redis.properties file.	Not available.
prestosql-connector-redshift	Change values in PrestoSQL's redshift.properties file.	Not available.
prestosql-connector-tpch	Change values in PrestoSQL's tpch.properties file.	Not available.
prestosql-connector-tpcds	Change values in PrestoSQL's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-log4j	Change values in kms-log4j .properties file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.

Classifications	Description	Reconfiguration Actions
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j	Change values in Spark's log4j.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restart Oozie.
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

Amazon EMR release 6.1.1

6.1.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, PrestoSQL, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
AWS SDK for Java	1.11.828	1.11.828	1.11.711	1.11.711
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Scala	2.12.10	2.12.10	2.12.10	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.0	1.11.0	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.5	2.2.5	2.2.3	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1	0.5.0-incubating-amzn-1
Hue	4.7.1	4.7.1	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.1.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Pig	0.17.0	0.17.0	-	-
Presto	0.232	0.232	0.230	0.230
Spark	3.0.0	3.0.0	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	-	-
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	338	338	-	-
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

6.1.1 release notes

This is a release to fix issues with Amazon EMR Scaling when it fails to scale up/scale down a cluster successfully or causes application failures.

Changes, Enhancements, and Resolved Issues

- Fixed an issue where scaling requests failed for a large, highly utilized cluster when Amazon EMR on-cluster daemons were running health checking activities, such as gathering YARN node state and HDFS node state. This was happening because on-cluster daemons were not able to communicate the health status data of a node to internal Amazon EMR components.
- Improved EMR on-cluster daemons to correctly track the node states when IP addresses are reused to improve reliability during scaling operations.
- SPARK-29683. Fixed an issue where job failures occurred during cluster scale-down as Spark was assuming all available nodes were deny-listed.
- YARN-9011. Fixed an issue where job failures occurred due to a race condition in YARN decommissioning when cluster tried to scale up or down.

- Fixed issue with step or job failures during cluster scaling by ensuring that the node states are always consistent between the Amazon EMR on-cluster daemons and YARN/HDFS.
- Fixed an issue where cluster operations such as scale down and step submission failed for Amazon EMR clusters enabled with Kerberos authentication. This was because the Amazon EMR on-cluster daemon did not renew the Kerberos ticket, which is required to securely communicate with HDFS/YARN running on the primary node.
- Newer Amazon EMR releases fix the issue with a lower "Max open files" limit on older AL2 in Amazon EMR. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later now include a permanent fix with a higher "Max open files" setting.
- HTTPS is now enabled by default for Amazon Linux repositories. If you are using an Amazon S3 VPCE policy to restrict access to specific buckets, you must add the new Amazon Linux bucket ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` to your policy (replace `$region` with the region where the endpoint is). For more information, see this topic in the AWS discussion forums. Announcement: Amazon Linux 2 now supports the ability to use HTTPS while connecting to package repositories .

6.1.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.3.0	Amazon SageMaker Spark SDK

Component	Version	Description
emr-ddb	4.14.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.1.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.14.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.0.0	EMR S3Select Connector
emrfs	2.42.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.11.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.

Component	Version	Description
hadoop-client	3.2.1-amzn-1.1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-1.1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	3.2.1-amzn-1.1	HDFS command-line client and library
hadoop-hdfs-namenode	3.2.1-amzn-1.1	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	3.2.1-amzn-1.1	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	3.2.1-amzn-1.1	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.2.1-amzn-1.1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.2.1-amzn-1.1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-1.1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.2.1-amzn-1.1	YARN service for allocatin g and managing cluster resources and distributed applications.

Component	Version	Description
hadoop-yarn-timeline-server	3.2.1-amzn-1.1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	2.2.5	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.2.5	Service for serving one or more HBase regions.
hbase-client	2.2.5	HBase command-line client.
hbase-rest-server	2.2.5	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	2.2.5	Service providing a Thrift endpoint to HBase.
hcatalog-client	3.1.2-amzn-2	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.2-amzn-2	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.2-amzn-2	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.2-amzn-2	Hive command line client.
hive-hbase	3.1.2-amzn-2	Hive-hbase client.

Component	Version	Description
hive-metastore-server	3.1.2-amzn-2	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	3.1.2-amzn-2	Service for accepting Hive queries as web requests.
hudi	0.5.2-incubating-amzn-2	Incremental processing framework to power data pipline at low latency and high efficiency.
hudi-presto	0.5.2-incubating-amzn-2	Bundle library for running Presto with Hudi.
hudi-prestosql	0.5.2-incubating-amzn-2	Bundle library for running PrestoSQL with Hudi.
hudi-spark	0.5.2-incubating-amzn-2	Bundle library for running Spark with Hudi.
hue-server	4.7.1	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.1.0	Multi-user server for Jupyter notebooks
livy-server	0.7.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server

Component	Version	Description
mxnet	1.6.0	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.64+	MariaDB database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.2.0	Oozie command-line client.
oozie-server	5.2.0	Service for accepting Oozie workflow requests.
opencv	4.3.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	The phoenix libraries for server and client
phoenix-query-server	5.0.0-HBase-2.0	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.232	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.232	Service for executing pieces of a query.

Component	Version	Description
presto-client	0.232	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
prestosql-coordinator	338	Service for accepting queries and managing query execution among prestosql-workers.
prestosql-worker	338	Service for executing pieces of a query.
prestosql-client	338	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
pig-client	0.17.0	Pig command-line client.
r	3.4.3	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.0.0-amzn-0.1	Spark command-line clients.
spark-history-server	3.0.0-amzn-0.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	3.0.0-amzn-0.1	In-memory execution engine for YARN.

Component	Version	Description
spark-yarn-slave	3.0.0-amzn-0.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.1.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.41+	Apache HTTP server.
zeppelin-server	0.9.0-preview1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

6.1.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-6.1.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.

Classifications	Description
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-env	Change values in the HDFS environment.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive	Amazon EMR-curated settings for Apache Hive.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.

Classifications	Description
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2- site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
hudi-env	Change values in the Hudi environment.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.

Classifications	Description
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.

Classifications	Description
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.

Classifications	Description
presto-connector-postgresql	Change values in Presto's postgresql.properties file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.
prestosql-log	Change values in Presto's log.properties file.
prestosql-config	Change values in Presto's config.properties file.
prestosql-password-authenticator	Change values in Presto's password-authenticator.properties file.
prestosql-env	Change values in Presto's presto-env.sh file.
prestosql-node	Change values in PrestoSQL's node.properties file.
prestosql-connector-blackhole	Change values in PrestoSQL's blackhole.properties file.
prestosql-connector-cassandra	Change values in PrestoSQL's cassandra.properties file.
prestosql-connector-hive	Change values in PrestoSQL's hive.properties file.

Classifications	Description
prestosql-connector-jmx	Change values in PrestoSQL's jmx.properties file.
prestosql-connector-kafka	Change values in PrestoSQL's kafka.properties file.
prestosql-connector-localfile	Change values in PrestoSQL's localfile.properti es file.
prestosql-connector-memory	Change values in PrestoSQL's memory.pr operties file.
prestosql-connector-mongodb	Change values in PrestoSQL's mongodb.p roperties file.
prestosql-connector-mysql	Change values in PrestoSQL's mysql.properties file.
prestosql-connector-postgresql	Change values in PrestoSQL's postgresq l.properties file.
prestosql-connector-raptor	Change values in PrestoSQL's raptor.pr operties file.
prestosql-connector-redis	Change values in PrestoSQL's redis.properties file.
prestosql-connector-redshift	Change values in PrestoSQL's redshift. properties file.
prestosql-connector-tpch	Change values in PrestoSQL's tpch.properties file.
prestosql-connector-tpcds	Change values in PrestoSQL's tpcds.properties file.
ranger-kms-dbks-site	Change values in dbks-site.xml file of Ranger KMS.

Classifications	Description
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.
ranger-kms-env	Change values in the Ranger KMS environment.
ranger-kms-log4j	Change values in kms-log4j.properties file of Ranger KMS.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.

Classifications	Description
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 6.1.0

6.1.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, JupyterHub, Livy, MXNet, Oozie, Phoenix, Pig, Presto, PrestoSQL, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
AWS SDK for Java	1.11.828	1.11.828	1.11.711	1.11.711
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.12.10	2.12.10	2.12.10	2.11.12

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.0	1.11.0	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.5	2.2.5	2.2.3	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1	0.5.0-incubating-amzn-1
Hue	4.7.1	4.7.1	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.1.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	-	-

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Presto	0.232	0.232	0.230	0.230
Spark	3.0.0	3.0.0	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	-	-
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	338	338	-	-
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

6.1.0 release notes

The following release notes include information for Amazon EMR release 6.1.0. Changes are relative to 6.0.0.

Initial release date: Sept 04, 2020

Last updated date: Oct 15, 2020

Supported applications

- AWS SDK for Java version 1.11.828
- Flink version 1.11.0
- Ganglia version 3.7.2
- Hadoop version 3.2.1-amzn-1
- HBase version 2.2.5
- HBase-operator-tools 1.0.0
- HCatalog version 3.1.2-amzn-0
- Hive version 3.1.2-amzn-1

- Hudi version 0.5.2-incubating
- Hue version 4.7.1
- JupyterHub version 1.1.0
- Livy version 0.7.0
- MXNet version 1.6.0
- Oozie version 5.2.0
- Phoenix version 5.0.0
- Presto version 0.232
- PrestoSQL version 338
- Spark version 3.0.0-amzn-0
- TensorFlow version 2.1.0
- Zeppelin version 0.9.0-preview1
- Zookeeper version 3.4.14
- Connectors and drivers: DynamoDB Connector 4.14.0

New features

- ARM instance types are supported starting with Amazon EMR version 5.30.0 and Amazon EMR version 6.1.0.
- M6g general purpose instance types are supported starting with Amazon EMR versions 6.1.0 and 5.30.0. For more information, see Supported Instance Types in the *Amazon EMR Management Guide*.
- The EC2 placement group feature is supported starting with Amazon EMR version 5.23.0 as an option for multiple primary node clusters. Currently, only primary node types are supported by the placement group feature, and the SPREAD strategy is applied to those primary nodes. The SPREAD strategy places a small group of instances across separate underlying hardware to guard against the loss of multiple primary nodes in the event of a hardware failure. For more information, see EMR Integration with EC2 Placement Group in the *Amazon EMR Management Guide*.
- Managed Scaling – With Amazon EMR version 6.1.0, you can enable Amazon EMR managed scaling to automatically increase or decrease the number of instances or units in your cluster based on workload. Amazon EMR continuously evaluates cluster metrics to make scaling decisions that optimize your clusters for cost and speed. Managed Scaling is also available on

Amazon EMR version 5.30.0 and later, except 6.0.0. For more information, see Scaling Cluster Resources in the *Amazon EMR Management Guide*.

- PrestoSQL version 338 is supported with EMR 6.1.0. For more information, see Presto.
  - PrestoSQL is supported on EMR 6.1.0 and later versions only, not on EMR 6.0.0 or EMR 5.x.
  - The application name, `Presto` continues to be used to install PrestoDB on clusters. To install PrestoSQL on clusters, use the application name `PrestoSQL`.
  - You can install either PrestoDB or PrestoSQL, but you cannot install both on a single cluster. If both PrestoDB and PrestoSQL are specified when attempting to create a cluster, a validation error occurs and the cluster creation request fails.
  - PrestoSQL is supported on both single-master and muti-master clusters. On multi-master clusters, an external Hive metastore is required to run PrestoSQL or PrestoDB. See Supported applications in an EMR cluster with multiple primary nodes.
- ECR auto authentication support on Apache Hadoop and Apache Spark with Docker: Spark users can use Docker images from Docker Hub and Amazon Elastic Container Registry (Amazon ECR) to define environment and library dependencies.

  Configure Docker and Run Spark Applications with Docker Using Amazon EMR 6.x.
- EMR supports Apache Hive ACID transactions: Amazon EMR 6.1.0 adds support for Hive ACID transactions so it complies with the ACID properties of a database. With this feature, you can run `INSERT, UPDATE, DELETE,` and `MERGE` operations in Hive managed tables with data in Amazon Simple Storage Service (Amazon S3). This is a key feature for use cases like streaming ingestion, data restatement, bulk updates using MERGE, and slowly changing dimensions. For more information, including configuration examples and use cases, see Amazon EMR supports Apache Hive ACID transactions.

Changes, enhancements, and resolved issues

- This is a release to fix issues with Amazon EMR Scaling when it fails to scale up/scale down a cluster successfully or causes application failures.
- Fixed an issue where scaling requests failed for a large, highly utilized cluster when Amazon EMR on-cluster daemons were running health checking activities, such as gathering YARN node state and HDFS node state. This was happening because on-cluster daemons were not able to communicate the health status data of a node to internal Amazon EMR components.
- Improved EMR on-cluster daemons to correctly track the node states when IP addresses are reused to improve reliability during scaling operations.

- SPARK-29683. Fixed an issue where job failures occurred during cluster scale-down as Spark was assuming all available nodes were deny-listed.
- YARN-9011. Fixed an issue where job failures occurred due to a race condition in YARN decommissioning when cluster tried to scale up or down.
- Fixed issue with step or job failures during cluster scaling by ensuring that the node states are always consistent between the Amazon EMR on-cluster daemons and YARN/HDFS.
- Fixed an issue where cluster operations such as scale down and step submission failed for Amazon EMR clusters enabled with Kerberos authentication. This was because the Amazon EMR on-cluster daemon did not renew the Kerberos ticket, which is required to securely communicate with HDFS/YARN running on the primary node.
- Newer Amazon EMR releases fix the issue with a lower "Max open files" limit on older AL2 in Amazon EMR. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later now include a permanent fix with a higher "Max open files" setting.
- Apache Flink is not supported on EMR 6.0.0, but it is supported on EMR 6.1.0 with Flink 1.11.0. This is the first version of Flink to officially support Hadoop 3. See Apache Flink 1.11.0 Release Announcement.
- Ganglia has been removed from default EMR 6.1.0 package bundles.

Known issues

- **Lower "Max open files" limit on older AL2 [fixed in newer releases].** Amazon EMR releases: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0, and emr-6.2.0 are based on older versions ofAmazon Linux 2 (AL2), which have a lower ulimit setting for "Max open files" when Amazon EMR clusters are created with the default AMI. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later include a permanent fix with a higher "Max open files" setting. Releases with the lower open file limit causes a "Too many open files" error when submitting Spark job. In the impacted releases, the Amazon EMR default AMI has a default ulimit setting of 4096 for "Max open files," which is lower than the 65536 file limit in the latestAmazon Linux 2 AMI. The lower ulimit setting for "Max open files" causes Spark job failure when the Spark driver and executor try to open more than 4096 files. To fix the issue, Amazon EMR has a bootstrap action (BA) script that adjusts the ulimit setting at cluster creation.

  If you are using an older Amazon EMR version that doesn't have the permanent fix for this issue, the following workaround lets you to explicitly set the instance-controller ulimit to a maximum of 65536 files.

Explicitly set a ulimit from the command line

1. Edit `/etc/systemd/system/instance-controller.service` to add the following parameters to Service section.

   `LimitNOFILE=65536`

   `LimitNPROC=65536`

2. Restart InstanceController

   `$ sudo systemctl daemon-reload`

   `$ sudo systemctl restart instance-controller`

Set a ulimit using bootstrap action (BA)

You can also use a bootstrap action (BA) script to configure the instance-controller ulimit to 65536 files at cluster creation.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- > **Important**
  >
  > Amazon EMR 6.1.0 and 6.2.0 include a performance issue that can critically affect all Hudi insert, upsert, and delete operations. If you plan to use Hudi with Amazon EMR 6.1.0 or 6.2.0, you should contact AWS support to obtain a patched Hudi RPM.
- If you set custom garbage collection configuration with `spark.driver.extraJavaOptions` and `spark.executor.extraJavaOptions`, this will result in driver/executor launch failure with EMR 6.1 due to conflicting garbage collection configuration. With EMR Release 6.1.0, you should specify custom Spark garbage collection configuration for drivers and executors with the properties `spark.driver.defaultJavaOptions` and `spark.executor.defaultJavaOptions` instead. Read more in Apache Spark Runtime Environment and Configuring Spark Garbage Collection on Amazon EMR 6.1.0.
- Using Pig with Oozie (and within Hue, since Hue uses Oozie actions to run Pig scripts), generates an error that a native-lzo library cannot be loaded. This error message is informational and does not block Pig from running.
- Hudi Concurrency Support: Currently Hudi doesn't support concurrent writes to a single Hudi table. In addition, Hudi rolls back any changes being done by in-progress writers before allowing a new writer to start. Concurrent writes can interfere with this mechanism and introduce race conditions, which can lead to data corruption. You should ensure that as part of your data processing workflow, there is only a single Hudi writer operating against a Hudi table at any time. Hudi does support multiple concurrent readers operating against the same Hudi table.
- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as hadoop user to the lead primary node of the EMR cluster with multiple primary nodes.
  - Run the following command to renew Kerberos ticket for hadoop user.

  ```
  kinit -kt <keytab_file> <principal>
  ```

Typically, the keytab file is located at `/etc/hadoop.keytab` and the principal is in the form of `hadoop/<hostname>@<REALM>`.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

- There is an issue in Amazon EMR 6.1.0 that affects clusters running Presto. After an extended period of time (days), the cluster may throw errors such as, "su: failed to execute /bin/bash: Resource temporarily unavailable" or "shell request failed on channel 0". This issue is caused by an internal Amazon EMR process (InstanceController) that is spawning too many Light Weight Processes (LWP), which eventually causes the Hadoop user to exceed their nproc limit. This prevents the user from opening additional processes. The solution for this issue is to upgrade to EMR 6.2.0.

6.1.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.3.0	Amazon SageMaker Spark SDK

Component	Version	Description
emr-ddb	4.14.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	3.1.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.14.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	2.0.0	EMR S3Select Connector
emrfs	2.42.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.11.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.

Component	Version	Description
hadoop-client	3.2.1-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	3.2.1-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	3.2.1-amzn-1	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	3.2.1-amzn-1	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	3.2.1-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.2.1-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.2.1-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resourcemanager	3.2.1-amzn-1	YARN service for allocating and managing cluster resources and distributed applications.

Component	Version	Description
hadoop-yarn-timeline-server	3.2.1-amzn-1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	2.2.5	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.2.5	Service for serving one or more HBase regions.
hbase-client	2.2.5	HBase command-line client.
hbase-rest-server	2.2.5	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	2.2.5	Service providing a Thrift endpoint to HBase.
hcatalog-client	3.1.2-amzn-2	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.2-amzn-2	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.2-amzn-2	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.2-amzn-2	Hive command line client.
hive-hbase	3.1.2-amzn-2	Hive-hbase client.

Component	Version	Description
hive-metastore-server	3.1.2-amzn-2	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	3.1.2-amzn-2	Service for accepting Hive queries as web requests.
hudi	0.5.2-incubating-amzn-2	Incremental processing framework to power data pipline at low latency and high efficiency.
hudi-presto	0.5.2-incubating-amzn-2	Bundle library for running Presto with Hudi.
hudi-prestosql	0.5.2-incubating-amzn-2	Bundle library for running PrestoSQL with Hudi.
hudi-spark	0.5.2-incubating-amzn-2	Bundle library for running Spark with Hudi.
hue-server	4.7.1	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.1.0	Multi-user server for Jupyter notebooks
livy-server	0.7.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server

Component	Version	Description
mxnet	1.6.0	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.64+	MariaDB database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.2.0	Oozie command-line client.
oozie-server	5.2.0	Service for accepting Oozie workflow requests.
opencv	4.3.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	The phoenix libraries for server and client
phoenix-query-server	5.0.0-HBase-2.0	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.232	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.232	Service for executing pieces of a query.

Component	Version	Description
presto-client	0.232	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
prestosql-coordinator	338	Service for accepting queries and managing query execution among prestosql-workers.
prestosql-worker	338	Service for executing pieces of a query.
prestosql-client	338	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
pig-client	0.17.0	Pig command-line client.
r	3.4.3	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.0.0-amzn-0	Spark command-line clients.
spark-history-server	3.0.0-amzn-0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	3.0.0-amzn-0	In-memory execution engine for YARN.

Component	Version	Description
spark-yarn-slave	3.0.0-amzn-0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.1.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.41+	Apache HTTP server.
zeppelin-server	0.9.0-preview1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

6.1.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-6.1.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.

Classifications	Description
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-env	Change values in the HDFS environment.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive	Amazon EMR-curated settings for Apache Hive.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.

Classifications	Description
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2- site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
hudi-env	Change values in the Hudi environment.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.

Classifications	Description
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.

Classifications	Description
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.

Classifications	Description
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.
prestosql-log	Change values in Presto's log.properties file.
prestosql-config	Change values in Presto's config.properties file.
prestosql-password-authenticator	Change values in Presto's password-authentic ator.properties file.
prestosql-env	Change values in Presto's presto-env.sh file.
prestosql-node	Change values in PrestoSQL's node.properties file.
prestosql-connector-blackhole	Change values in PrestoSQL's blackhole .properties file.
prestosql-connector-cassandra	Change values in PrestoSQL's cassandra .properties file.
prestosql-connector-hive	Change values in PrestoSQL's hive.properties file.

Classifications	Description
prestosql-connector-jmx	Change values in PrestoSQL's jmx.properties file.
prestosql-connector-kafka	Change values in PrestoSQL's kafka.properties file.
prestosql-connector-localfile	Change values in PrestoSQL's localfile.properti es file.
prestosql-connector-memory	Change values in PrestoSQL's memory.pr operties file.
prestosql-connector-mongodb	Change values in PrestoSQL's mongodb.p roperties file.
prestosql-connector-mysql	Change values in PrestoSQL's mysql.properties file.
prestosql-connector-postgresql	Change values in PrestoSQL's postgresq l.properties file.
prestosql-connector-raptor	Change values in PrestoSQL's raptor.pr operties file.
prestosql-connector-redis	Change values in PrestoSQL's redis.properties file.
prestosql-connector-redshift	Change values in PrestoSQL's redshift. properties file.
prestosql-connector-tpch	Change values in PrestoSQL's tpch.properties file.
prestosql-connector-tpcds	Change values in PrestoSQL's tpcds.properties file.
ranger-kms-dbks-site	Change values in dbks-site.xml file of Ranger KMS.

Classifications	Description
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.
ranger-kms-env	Change values in the Ranger KMS environme nt.
ranger-kms-log4j	Change values in kms-log4j.properties file of Ranger KMS.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.

Classifications	Description
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 6.0.1

6.0.1 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, JupyterHub, Livy, MXNet, Oozie, Phoenix, Presto, Spark, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
AWS SDK for Java	1.11.828	1.11.828	1.11.711	1.11.711
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.12.10	2.12.10	2.12.10	2.11.12

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.0	1.11.0	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.5	2.2.5	2.2.3	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1	0.5.0-incubating-amzn-1
Hue	4.7.1	4.7.1	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.1.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	-	-

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Presto	0.232	0.232	0.230	0.230
Spark	3.0.0	3.0.0	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	-	-
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	338	338	-	-
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

6.0.1 release notes

This is a release to fix issues with Amazon EMR Scaling when it fails to scale up/scale down a cluster successfully or causes application failures.

Changes, Enhancements, and Resolved Issues

- Fixed an issue where scaling requests failed for a large, highly utilized cluster when Amazon EMR on-cluster daemons were running health checking activities, such as gathering YARN node state and HDFS node state. This was happening because on-cluster daemons were not able to communicate the health status data of a node to internal Amazon EMR components.
- Improved EMR on-cluster daemons to correctly track the node states when IP addresses are reused to improve reliability during scaling operations.
- SPARK-29683. Fixed an issue where job failures occurred during cluster scale-down as Spark was assuming all available nodes were deny-listed.
- YARN-9011. Fixed an issue where job failures occurred due to a race condition in YARN decommissioning when cluster tried to scale up or down.
- Fixed issue with step or job failures during cluster scaling by ensuring that the node states are always consistent between the Amazon EMR on-cluster daemons and YARN/HDFS.

- Fixed an issue where cluster operations such as scale down and step submission failed for Amazon EMR clusters enabled with Kerberos authentication. This was because the Amazon EMR on-cluster daemon did not renew the Kerberos ticket, which is required to securely communicate with HDFS/YARN running on the primary node.
- Newer Amazon EMR releases fix the issue with a lower "Max open files" limit on older AL2 in Amazon EMR. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later now include a permanent fix with a higher "Max open files" setting.
- HTTPS is now enabled by default for Amazon Linux repositories. If you are using an Amazon S3 VPCE policy to restrict access to specific buckets, you must add the new Amazon Linux bucket ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` to your policy (replace `$region` with the region where the endpoint is). For more information, see this topic in the AWS discussion forums. [Announcement: Amazon Linux 2 now supports the ability to use HTTPS while connecting to package repositories](#).

6.0.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.2.6	Amazon SageMaker Spark SDK
emr-ddb	4.14.0	Amazon DynamoDB connector for Hadoop ecosystem applications.

Component	Version	Description
emr-goodies	3.0.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.14.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.5.0	EMR S3Select Connector
emrfs	2.39.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	3.2.1-amzn-0.1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-0.1	HDFS node-level service for storing blocks.

Component	Version	Description
hadoop-hdfs-library	3.2.1-amzn-0.1	HDFS command-line client and library
hadoop-hdfs-namenode	3.2.1-amzn-0.1	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	3.2.1-amzn-0.1	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	3.2.1-amzn-0.1	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.2.1-amzn-0.1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.2.1-amzn-0.1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-0.1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.2.1-amzn-0.1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.2.1-amzn-0.1	Service for retrieving current and historical information for YARN applications.

Component	Version	Description
hbase-hmaster	2.2.3	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.2.3	Service for serving one or more HBase regions.
hbase-client	2.2.3	HBase command-line client.
hbase-rest-server	2.2.3	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	2.2.3	Service providing a Thrift endpoint to HBase.
hcatalog-client	3.1.2-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.2-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.2-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.2-amzn-0	Hive command line client.
hive-hbase	3.1.2-amzn-0	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.

Component	Version	Description
hive-server2	3.1.2-amzn-0	Service for accepting Hive queries as web requests.
hudi	0.5.0-incubating-amzn-1	Incremental processing framework to power data pipline at low latency and high efficiency.
hudi-presto	0.5.0-incubating-amzn-1	Bundle library for running Presto with Hudi.
hue-server	4.4.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.0.0	Multi-user server for Jupyter notebooks
livy-server	0.6.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.5.1	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.64+	MariaDB database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.1.0	Oozie command-line client.
oozie-server	5.1.0	Service for accepting Oozie workflow requests.

Component	Version	Description
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	The phoenix libraries for server and client
phoenix-query-server	5.0.0-HBase-2.0	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.230	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.230	Service for executing pieces of a query.
presto-client	0.230	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
r	3.4.3	The R Project for Statistical Computing
spark-client	2.4.4	Spark command-line clients.
spark-history-server	2.4.4	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.4	In-memory execution engine for YARN.

Component	Version	Description
spark-yarn-slave	2.4.4	Apache Spark libraries needed by YARN slaves.
tensorflow	1.14.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.41+	Apache HTTP server.
zeppelin-server	0.9.0-SNAPSHOT	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

6.0.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-6.0.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.

Classifications	Description
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-env	Change values in the HDFS environment.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive	Amazon EMR-curated settings for Apache Hive.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.

Classifications	Description
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.

Classifications	Description
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.

Classifications	Description
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.
ranger-kms-dbks-site	Change values in dbks-site.xml file of Ranger KMS.

Classifications	Description
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.
ranger-kms-env	Change values in the Ranger KMS environme nt.
ranger-kms-log4j	Change values in kms-log4j.properties file of Ranger KMS.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.
recordserver-env	Change values in the EMR RecordServer environment.
recordserver-conf	Change values in EMR RecordServer's erver.pro perties file.
recordserver-log4j	Change values in EMR RecordServer's log4j.pro perties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
tez-site	Change values in Tez's tez-site.xml file.

Classifications	Description
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 6.0.0

6.0.0 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, JupyterHub, Livy, MXNet, Oozie, Phoenix, Presto, Spark, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
AWS SDK for Java	1.11.828	1.11.828	1.11.711	1.11.711

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.12.10	2.12.10	2.12.10	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.0	1.11.0	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.5	2.2.5	2.2.3	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1	0.5.0-incubating-amzn-1
Hue	4.7.1	4.7.1	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.1.0	5.1.0

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	-	-
Presto	0.232	0.232	0.230	0.230
Spark	3.0.0	3.0.0	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	-	-
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	338	338	-	-
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

6.0.0 release notes

The following release notes include information for Amazon EMR release 6.0.0.

Initial release date: March 10, 2020

Supported applications

- AWS SDK for Java version 1.11.711
- Ganglia version 3.7.2
- Hadoop version 3.2.1
- HBase version 2.2.3
- HCatalog version 3.1.2
- Hive version 3.1.2
- Hudi version 0.5.0-incubating

- Hue version 4.4.0
- JupyterHub version 1.0.0
- Livy version 0.6.0
- MXNet version 1.5.1
- Oozie version 5.1.0
- Phoenix version 5.0.0
- Presto version 0.230
- Spark version 2.4.4
- TensorFlow version 1.14.0
- Zeppelin version 0.9.0-SNAPSHOT
- Zookeeper version 3.4.14
- Connectors and drivers: DynamoDB Connector 4.14.0

> **Note**
>
> Flink, Sqoop, Pig, and Mahout are not available in Amazon EMR version 6.0.0.

New features

- YARN Docker Runtime Support - YARN applications, such as Spark jobs, can now run in the context of a Docker container. This allows you to easily define dependencies in a Docker image without the need to install custom libraries on your Amazon EMR cluster. For more information, see Configure Docker Integration and Run Spark applications with Docker using Amazon EMR 6.0.0.
- Hive LLAP Support - Hive now supports the LLAP execution mode for improved query performance. For more information, see Using Hive LLAP.

Changes, enhancements, and resolved issues

- This is a release to fix issues with Amazon EMR Scaling when it fails to scale up/scale down a cluster successfully or causes application failures.
- Fixed an issue where scaling requests failed for a large, highly utilized cluster when Amazon EMR on-cluster daemons were running health checking activities, such as gathering YARN node

state and HDFS node state. This was happening because on-cluster daemons were not able to communicate the health status data of a node to internal Amazon EMR components.

- Improved EMR on-cluster daemons to correctly track the node states when IP addresses are reused to improve reliability during scaling operations.
- SPARK-29683. Fixed an issue where job failures occurred during cluster scale-down as Spark was assuming all available nodes were deny-listed.
- YARN-9011. Fixed an issue where job failures occurred due to a race condition in YARN decommissioning when cluster tried to scale up or down.
- Fixed issue with step or job failures during cluster scaling by ensuring that the node states are always consistent between the Amazon EMR on-cluster daemons and YARN/HDFS.
- Fixed an issue where cluster operations such as scale down and step submission failed for Amazon EMR clusters enabled with Kerberos authentication. This was because the Amazon EMR on-cluster daemon did not renew the Kerberos ticket, which is required to securely communicate with HDFS/YARN running on the primary node.
- Newer Amazon EMR releases fix the issue with a lower "Max open files" limit on older AL2 in Amazon EMR. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later now include a permanent fix with a higher "Max open files" setting.
- Amazon Linux
  - Amazon Linux 2 is the operating system for the EMR 6.x release series.
  - `systemd` is used for service management instead of `upstart` used inAmazon Linux 1.
- Java Development Kit (JDK)
  - Corretto JDK 8 is the default JDK for the EMR 6.x release series.
- Scala
  - Scala 2.12 is used with Apache Spark and Apache Livy.
- Python 3
  - Python 3 is now the default version of Python in EMR.
- YARN node labels
  - Beginning with Amazon EMR 6.x release series, the YARN node labels feature is disabled by default. The application master processes can run on both core and task nodes by default. You can enable the YARN node labels feature by configuring following properties: `yarn.node-labels.enabled` and `yarn.node-labels.am.default-node-label-expression`. For more information, see Understanding Primary, Core, and Task Nodes.

Known issues

- **Lower "Max open files" limit on older AL2 [fixed in newer releases].** Amazon EMR releases: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0, and emr-6.2.0 are based on older versions ofAmazon Linux 2 (AL2), which have a lower ulimit setting for "Max open files" when Amazon EMR clusters are created with the default AMI. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later include a permanent fix with a higher "Max open files" setting. Releases with the lower open file limit causes a "Too many open files" error when submitting Spark job. In the impacted releases, the Amazon EMR default AMI has a default ulimit setting of 4096 for "Max open files," which is lower than the 65536 file limit in the latestAmazon Linux 2 AMI. The lower ulimit setting for "Max open files" causes Spark job failure when the Spark driver and executor try to open more than 4096 files. To fix the issue, Amazon EMR has a bootstrap action (BA) script that adjusts the ulimit setting at cluster creation.

  If you are using an older Amazon EMR version that doesn't have the permanent fix for this issue, the following workaround lets you to explicitly set the instance-controller ulimit to a maximum of 65536 files.

  Explicitly set a ulimit from the command line

  1. Edit `/etc/systemd/system/instance-controller.service` to add the following parameters to Service section.

     `LimitNOFILE=65536`

     `LimitNPROC=65536`
  2. Restart InstanceController

     `$ sudo systemctl daemon-reload`

     `$ sudo systemctl restart instance-controller`

  Set a ulimit using bootstrap action (BA)

  You can also use a bootstrap action (BA) script to configure the instance-controller ulimit to 65536 files at cluster creation.

```
#!/bin/bash
for user in hadoop spark hive; do
```

```
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- Spark interactive shell, including PySpark, SparkR, and spark-shell, does not support using Docker with additional libraries.
- To use Python 3 with Amazon EMR version 6.0.0, you must add PATH to `yarn.nodemanager.env-whitelist`.
- The Live Long and Process (LLAP) functionality is not supported when you use the AWS Glue Data Catalog as the metastore for Hive.
- When using Amazon EMR 6.0.0 with Spark and Docker integration, you need to configure the instances in your cluster with the same instance type and the same amount of EBS volumes to avoid failure when submitting a Spark job with Docker runtime.
- In Amazon EMR 6.0.0, HBase on Amazon S3 storage mode is impacted by the HBASE-24286. issue. HBase master cannot initialize when the cluster is created using existing S3 data.
- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as hadoop user to the lead primary node of the EMR cluster with multiple primary nodes.

- Run the following command to renew Kerberos ticket for hadoop user.

```
kinit -kt <keytab_file> <principal>
```

Typically, the keytab file is located at /etc/hadoop.keytab and the principal is in the form of hadoop/<hostname>@<REALM>.

> **Note**
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

6.0.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with emr or aws. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
aws-sagemaker-spark-sdk	1.2.6	Amazon SageMaker Spark SDK
emr-ddb	4.14.0	Amazon DynamoDB connector for Hadoop ecosystem applications.

Component	Version	Description
emr-goodies	3.0.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.14.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.5.0	EMR S3Select Connector
emrfs	2.39.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	3.2.1-amzn-0	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-0	HDFS node-level service for storing blocks.

Component	Version	Description
hadoop-hdfs-library	3.2.1-amzn-0	HDFS command-line client and library
hadoop-hdfs-namenode	3.2.1-amzn-0	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	3.2.1-amzn-0	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	3.2.1-amzn-0	HTTP endpoint for HDFS operations.
hadoop-kms-server	3.2.1-amzn-0	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	3.2.1-amzn-0	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-0	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	3.2.1-amzn-0	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	3.2.1-amzn-0	Service for retrieving current and historical information for YARN applications.

Component	Version	Description
hbase-hmaster	2.2.3	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	2.2.3	Service for serving one or more HBase regions.
hbase-client	2.2.3	HBase command-line client.
hbase-rest-server	2.2.3	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	2.2.3	Service providing a Thrift endpoint to HBase.
hcatalog-client	3.1.2-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	3.1.2-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	3.1.2-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	3.1.2-amzn-0	Hive command line client.
hive-hbase	3.1.2-amzn-0	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.

Component	Version	Description
hive-server2	3.1.2-amzn-0	Service for accepting Hive queries as web requests.
hudi	0.5.0-incubating-amzn-1	Incremental processing framework to power data pipline at low latency and high efficiency.
hudi-presto	0.5.0-incubating-amzn-1	Bundle library for running Presto with Hudi.
hue-server	4.4.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.0.0	Multi-user server for Jupyter notebooks
livy-server	0.6.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mxnet	1.5.1	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.64+	MariaDB database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.1.0	Oozie command-line client.
oozie-server	5.1.0	Service for accepting Oozie workflow requests.

Component	Version	Description
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	The phoenix libraries for server and client
phoenix-query-server	5.0.0-HBase-2.0	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.230	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.230	Service for executing pieces of a query.
presto-client	0.230	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
r	3.4.3	The R Project for Statistical Computing
spark-client	2.4.4	Spark command-line clients.
spark-history-server	2.4.4	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.4	In-memory execution engine for YARN.

Component	Version	Description
spark-yarn-slave	2.4.4	Apache Spark libraries needed by YARN slaves.
tensorflow	1.14.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.41+	Apache HTTP server.
zeppelin-server	0.9.0-SNAPSHOT	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

6.0.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see [Configure applications](#).

emr-6.0.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.

Classifications	Description
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-env	Change values in the HDFS environment.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive	Amazon EMR-curated settings for Apache Hive.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.

Classifications	Description
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.

Classifications	Description
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.

Classifications	Description
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.
ranger-kms-dbks-site	Change values in dbks-site.xml file of Ranger KMS.

Classifications	Description
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.
ranger-kms-env	Change values in the Ranger KMS environme nt.
ranger-kms-log4j	Change values in kms-log4j.properties file of Ranger KMS.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.
recordserver-env	Change values in the EMR RecordServer environment.
recordserver-conf	Change values in EMR RecordServer's erver.pro perties file.
recordserver-log4j	Change values in EMR RecordServer's log4j.pro perties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
tez-site	Change values in Tez's tez-site.xml file.

Classifications	Description
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR 5.x release versions

This section contains application versions, release notes, component versions, and configuration classifications available in each Amazon EMR 5.x release version.

When you launch a cluster, you can choose from multiple releases of Amazon EMR. This allows you to test and use application versions that fit your compatibility requirements. You specify the release number with the *release label*. Release labels are in the form `emr-x.x.x`. For example, `emr-7.2.0`.

New Amazon EMR releases are made available in different Regions over a period of several days, beginning with the first Region on the initial release date. The latest release version may not be available in your Region during this period.

For a comprehensive table of application versions in every Amazon EMR 5.x release, see Application versions in Amazon EMR 5.x releases.

Topics

Application versions in Amazon EMR 5.x releases

For a comprehensive table that lists the application versions available in each Amazon EMR 5.x release, open Application versions in Amazon EMR 5.x releases in your browser.

Amazon EMR release 5.36.2

5.36.2 application versions

The following applications are supported in this release: Delta, Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, Iceberg, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.36.2	emr-5.36.1	emr-5.36.0	emr-5.35.0
AWS SDK for Java	1.12.206	1.12.206	1.12.206	1.12.159
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.14.2	1.14.2	1.14.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13

	emr-5.36.2	emr-5.36.1	emr-5.36.0	emr-5.35.0
HCatalog	2.3.9	2.3.9	2.3.9	2.3.9
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.9	2.3.9	2.3.9	2.3.9
Hudi	0.10.1-amzn-1	0.10.1-amzn-1	0.10.1-amzn-1	0.9.0-amzn-2
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.6.0	2.6.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.8.0	1.8.0	1.8.0	1.8.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.267	0.267	0.267	0.266
Spark	2.4.8	2.4.8	2.4.8	2.4.8
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.36.2	emr-5.36.1	emr-5.36.0	emr-5.35.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.10.0	0.10.0	0.10.0	0.10.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

5.36.2 release notes

The following release notes include information for Amazon EMR release 5.36.2. Changes are relative to 5.36.1. For information on the release timeline, see the change log.

Changes, enhancements, and resolved issues

- This releases improves cluster scale-down logic so that Amazon EMR doesn't scale-down core nodes below the HDFS replication factor setting for the cluster. This improvement fulfills data redundancy requirements, and reduces the chance that a scaling operation might stall.
- This release adds a new retry mechanism to the cluster scaling workflow for that run Presto or Trino. This improvement reduces the risk that cluster resize runs indefinitely due to a single failed resize operation. It also improves cluster utilization, because your cluster scales up and down faster.
- Fixes an issue where cluster scale-down operations might stall while Amazon EMR gracefully decommissions a core node and it turns unhealthy before it is fully decommissioned.
- Improves the stability of a node in a high-availability cluster with multiple primary nodes when Amazon EMR restarts a single node.
- Optimizes log management with Amazon EMR running on Amazon EC2. As a result, you might see a slight reduction in storage costs for your cluster logs.
- Improves the management of ZooKeeper transaction log files that are maintained on primary nodes to minimize scenarios where the log files grow out of bounds and interrupt cluster operations.
- Fixes a rare bug which can cause a high-availability cluster with multiple primary nodes to fail due to not being able to communicate with the Yarn ResourceManager.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2

release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240816.0	4.14.350	August 21, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 809.0	4.14.349	August 20, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 719.0	4.14.348	July 25, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsRelease Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 503.0	4.14.343	xxxxxx, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

5.36.2 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the

most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.16.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.5.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.21.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.7.0	EMR S3Select Connector
emrfs	2.51.0	Amazon S3 connector for Hadoop ecosystem applicati ons.

Component	Version	Description
flink-client	1.14.2	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.14.2	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.10.1-amzn-4	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.10.1-amzn-4	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.10.1-amzn-4	HDFS command-line client and library
hadoop-hdfs-namenode	2.10.1-amzn-4	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	2.10.1-amzn-4	HDFS service for managing the Hadoop filesystem journal on HA clusters.

Component	Version	Description
hadoop-httpfs-server	2.10.1-amzn-4	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.10.1-amzn-4	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.10.1-amzn-4	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.10.1-amzn-4	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.10.1-amzn-4	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.10.1-amzn-4	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.13	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.13	Service for serving one or more HBase regions.
hbase-client	1.4.13	HBase command-line client.
hbase-rest-server	1.4.13	Service providing a RESTful HTTP endpoint for HBase.

Component	Version	Description
hbase-thrift-server	1.4.13	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.9-amzn-2	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.9-amzn-2	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.9-amzn-2	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.9-amzn-2	Hive command line client.
hive-hbase	2.3.9-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.9-amzn-2	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.9-amzn-2	Service for accepting Hive queries as web requests.
hudi	0.10.1-amzn-1	Incremental processing framework to power data pipeline at low latency and high efficiency.
hudi-spark	0.10.1-amzn-1	Bundle library for running Spark with Hudi.

Component	Version	Description
hudi-presto	0.10.1-amzn-1	Bundle library for running Presto with Hudi.
hue-server	4.10.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.4.1	Multi-user server for Jupyter notebooks
livy-server	0.7.1-incubating	REST interface for interacting with Apache Spark
nginx	1.13.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.8.0	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68	MySQL database server.
nvidia-cuda	11.0.194	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	The phoenix libraries for server and client

Component	Version	Description
phoenix-query-server	4.14.3-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.267-amzn-1	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.267-amzn-1	Service for executing pieces of a query.
presto-client	0.267-amzn-1	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.8-amzn-2	Spark command-line clients.
spark-history-server	2.4.8-amzn-2	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.8-amzn-2	In-memory execution engine for YARN.

Component	Version	Description
spark-yarn-slave	2.4.8-amzn-2	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.4.1	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.10.0	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

5.36.2 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-5.36.2 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode,

Classifications	Description	Reconfiguration Actions
		SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-session.properties settings for Kubernetes/Yarn session.	Not available.
flink-log4j-cli	Change Flink log4j-cli.properties settings.	Restarts Flink history server.

Classifications	Description	Reconfiguration Actions
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts PhoenixQu eryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.

Classifications	Description	Reconfiguration Actions
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	Should not be reconfigured.
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classifications	Description	Reconfiguration Actions
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.
hive-env	Change values in the Hive environment.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.

Classifications	Description	Reconfiguration Actions
hudi-env	Change values in the Hudi environment.	Not available.
hudi-defaults	Change values in Hudi's hudi-defaults.conf file.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j	Change Livy log4j.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
oozie-env	Change values in Oozie's environment.	Restarts Oozie.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hb ase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServe r.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.pro perties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server.
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server.

Classifications	Description	Reconfiguration Actions
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server.
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server.
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-log4j	Change values in kms-log4j .properties file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
recordserver-env	Change values in the EMR RecordServer environment.	Restarts EMR record server.

Classifications	Description	Reconfiguration Actions
recordserver-conf	Change values in EMR RecordServer's erver.pro perties file.	Restarts EMR record server.
recordserver-log4j	Change values in EMR RecordServer's log4j.pro perties file.	Restarts EMR record server.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j	Change values in Spark's log4j.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restarts Oozie.

Classifications	Description	Reconfiguration Actions
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts MapReduce -HistoryServer.
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Livy Server and MapReduce-HistoryS erver.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

5.36.2 change log

Change log for 5.36.2 release and release notes

Date	Event	Description
2024-05-28	Deployment complete	Amazon EMR 5.36.2 fully deployed to all supported Regions

Date	Event	Description
2024-05-28	Docs publication	Amazon EMR 5.36.2 release notes first published
2024-05-20	Initial release	Amazon EMR 5.36.2 first deployed to limited commercial Regions

Amazon EMR release 5.36.1

5.36.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, Iceberg, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.36.1	emr-5.36.0	emr-5.35.0	emr-5.34.0
AWS SDK for Java	1.12.206	1.12.206	1.12.159	1.11.970
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.11.12	2.11.12	2.11.12	2.11.12

	emr-5.36.1	emr-5.36.0	emr-5.35.0	emr-5.34.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.14.2	1.14.2	1.13.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.9	2.3.9	2.3.9	2.3.8
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.9	2.3.9	2.3.9	2.3.8
Hudi	0.10.1-amzn-1	0.10.1-amzn-1	0.9.0-amzn-2	0.9.0-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.9.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.6.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.8.0	1.8.0	1.8.0	1.8.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-5.36.1	emr-5.36.0	emr-5.35.0	emr-5.34.0
Presto	0.267	0.267	0.266	0.261
Spark	2.4.8	2.4.8	2.4.8	2.4.8
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.10.0	0.10.0	0.10.0	0.10.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

5.36.1 release notes

The following release notes include information for Amazon EMR release 5.36.1. Changes are relative to 5.36.0. For information on the release timeline, see the change log.

Changes, enhancements, and resolved issues

- Amazon EMR release 5.36.1 adds support for archiving logs to Amazon S3 during cluster scale-down. In previous 5.x releases, you could only archive log files to Amazon S3 during cluster termination. This improvement ensures that log files generated on the cluster persist on Amazon S3 even after the node is terminated. For more information, see Configure cluster logging and debugging.
- The 5.36.1 release improves the on-cluster log management daemon to monitor additional log folders in your EMR cluster. This improvement minimizes disk over-utilization scenarios.
- The 5.36.1 release automatically restarts the on-cluster log management daemon when it stops. This improvement reduces the risk for nodes to appear unhealthy due to disk over-utilization.
- The 5.36.1 release fixes an issue where Amazon EMR daemons on the primary node would maintain stale metadata for terminated instances in the cluster. Maintaining stale data might

cause on-cluster CPU and memory usage to grow without bounds, and ultimately cause cluster failures.

- For clusters that are launched with multiple primary nodes, the 5.36.1 release fixes an issue where an Amazon EC2 hardware failure on one of the primary nodes could cause a second primary node to fail and render your cluster unstable.
- For clusters that are configured with in-transit encryption, Managed Scaling is now *Spark shuffle data* aware. Spark shuffle data is data that Spark redistributes across partitions to perform specific operations. During scale down, Managed Scaling ignores the instances with shuffle data. This prevents job re-attempts and re-computations, which are costly for price and performance. For more information on shuffle operations, see the Spark Programming Guide.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

OsRelea Label (Amazoı Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 816.0	4.14.350	August 21, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
			(Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 809.0	4.14.349	August 20, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 719.0	4.14.348	July 25, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 223.0	4.14.336	March 8, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 131.0	4.14.336	February 14, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240124.0	4.14.336	February 7, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240109.0	4.14.334	January 24, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 218.0	4.14.330	January 2, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 206.0	4.14.330	December 22, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023116.0	4.14.328	December 11, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20231101.0	4.14.327	November 16, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazoı Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 020.1	4.14.326	November 7, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20231012.1	4.14.326	October 26, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023926.0	4.14.322	October 19, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230906.0	4.14.322	October 4, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230822.0	4.14.322	August 30, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 808.0	4.14.320	August 24, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230727.0	4.14.320	August 14, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230719.0	4.14.320	August 2, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 628.0	4.14.318	July 12, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 612.0	4.14.314	June 23, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230404.1	4.14.311	April 18, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

5.36.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.16.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.5.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.21.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.7.0	EMR S3Select Connector
emrfs	2.51.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.14.2	Apache Flink command line client scripts and applications.

Component	Version	Description
flink-jobmanager-config	1.14.2	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.10.1-amzn-4	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.10.1-amzn-4	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.10.1-amzn-4	HDFS command-line client and library
hadoop-hdfs-namenode	2.10.1-amzn-4	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	2.10.1-amzn-4	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.10.1-amzn-4	HTTP endpoint for HDFS operations.

Component	Version	Description
hadoop-kms-server	2.10.1-amzn-4	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.10.1-amzn-4	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.10.1-amzn-4	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.10.1-amzn-4	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.10.1-amzn-4	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.13	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.13	Service for serving one or more HBase regions.
hbase-client	1.4.13	HBase command-line client.
hbase-rest-server	1.4.13	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.13	Service providing a Thrift endpoint to HBase.

Component	Version	Description
hcatalog-client	2.3.9-amzn-2	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.9-amzn-2	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.9-amzn-2	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.9-amzn-2	Hive command line client.
hive-hbase	2.3.9-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.9-amzn-2	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.9-amzn-2	Service for accepting Hive queries as web requests.
hudi	0.10.1-amzn-1	Incremental processing framework to power data pipeline at low latency and high efficiency.
hudi-spark	0.10.1-amzn-1	Bundle library for running Spark with Hudi.
hudi-presto	0.10.1-amzn-1	Bundle library for running Presto with Hudi.

Component	Version	Description
hue-server	4.10.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.4.1	Multi-user server for Jupyter notebooks
livy-server	0.7.1-incubating	REST interface for interacting with Apache Spark
nginx	1.13.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.8.0	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68	MySQL database server.
nvidia-cuda	11.0.194	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	The phoenix libraries for server and client

Component	Version	Description
phoenix-query-server	4.14.3-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.267-amzn-1	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.267-amzn-1	Service for executing pieces of a query.
presto-client	0.267-amzn-1	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.8-amzn-2	Spark command-line clients.
spark-history-server	2.4.8-amzn-2	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.8-amzn-2	In-memory execution engine for YARN.

Component	Version	Description
spark-yarn-slave	2.4.8-amzn-2	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.4.1	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.10.0	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

5.36.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-5.36.1 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode,

Classifications	Description	Reconfiguration Actions
		SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-ses sion.properties settings for Kubernetes/Yarn session.	Not available.
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Restarts Flink history server.

Classifications	Description	Reconfiguration Actions
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.

Classifications	Description	Reconfiguration Actions
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	Should not be reconfigured.
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classifications	Description	Reconfiguration Actions
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.
hive-env	Change values in the Hive environment.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.

Classifications	Description	Reconfiguration Actions
hudi-env	Change values in the Hudi environment.	Not available.
hudi-defaults	Change values in Hudi's hudi-defaults.conf file.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j	Change Livy log4j.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
oozie-env	Change values in Oozie's environment.	Restarts Oozie.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hb ase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServe r.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.pro perties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server.
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server.

Classifications	Description	Reconfiguration Actions
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server.
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server.
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-log4j	Change values in kms-log4j .properties file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
recordserver-env	Change values in the EMR RecordServer environment.	Restarts EMR record server.

Classifications	Description	Reconfiguration Actions
recordserver-conf	Change values in EMR RecordServer's erver.pro perties file.	Restarts EMR record server.
recordserver-log4j	Change values in EMR RecordServer's log4j.pro perties file.	Restarts EMR record server.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j	Change values in Spark's log4j.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restarts Oozie.

Classifications	Description	Reconfiguration Actions
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts MapReduce -HistoryServer.
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Livy Server and MapReduce-HistoryS erver.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

5.36.1 change log

Change log for 5.36.1 release and release notes

Date	Event	Description
2023-07-26	Update	New OS release labels `2.0.20230612.0` and `2.0.20230628.0` .

Date	Event	Description
2023-05-25	Deployment complete	Amazon EMR 5.36.1 fully deployed to all supported Regions
2023-05-09	Docs publication	Amazon EMR 5.36.1 release notes first published
2023-05-04	Initial release	Amazon EMR 5.36.1 first deployed to limited commercial Regions

Amazon EMR release 5.36.0

5.36.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, Iceberg, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.36.0	emr-5.35.0	emr-5.34.0	emr-5.33.1
AWS SDK for Java	1.12.206	1.12.159	1.11.970	1.11.970
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.14.2	1.13.1	1.12.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.9	2.3.9	2.3.8	2.3.7
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.9	2.3.9	2.3.8	2.3.7
Hudi	0.10.1-amzn-1	0.9.0-amzn-2	0.9.0-amzn-0	0.7.0-amzn-1
Hue	4.10.0	4.10.0	4.9.0	4.9.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.2.2
Livy	0.7.1	0.7.1	0.7.1	0.7.0
MXNet	1.8.0	1.8.0	1.8.0	1.7.0

	emr-5.36.0	emr-5.35.0	emr-5.34.0	emr-5.33.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.1	5.2.1	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.267	0.266	0.261	0.245.1
Spark	2.4.8	2.4.8	2.4.8	2.4.7
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.10.0	0.10.0	0.10.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

5.36.0 release notes

The following release notes include information for Amazon EMR release 5.36.0. Changes are relative to 5.35.0.

Initial release date: June 15, 2022

New Features

- Amazon EMR release 5.36.0 adds support for data definition language (DDL) with Apache Spark on Apache Ranger enabled clusters. This allows you to use Apache Ranger for managing access for operations like creating, altering and dropping databases and tables from an Amazon EMR cluster.

- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

Note

This release no longer gets automatic AMI updates since it has been succeeded by 1 more more patch releases. The patch release is denoted by the number after the second decimal point (`6.8.1`). To see if you're using the latest patch release, check the available releases in the *Release Guide*, or check the **Amazon EMR release** dropdown when you create a cluster in the console, or use the `ListReleaseLabels` API or `list-release-labels` CLI action. To get updates about new releases, subscribe to the RSS feed on the What's new? page.

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
			(Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230504.1	4.14.313	May 16, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230418.0	4.14.311	May 3, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230404.1	4.14.311	April 18, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)
2.0.20230404.0	4.14.311	April 10, 2023	US East (N. Virginia), Europe (Paris)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230320.0	4.14.309	March 30, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230307.0	4.14.305	March 15, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230207.0	4.14.304	February 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230119.1	4.14.301	February 3, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelease Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20221210.1	4.14.301	December 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20221103.3	4.14.296	December 5, 2022	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelease Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022004.0	4.14.294	November 2, 2022	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 912.1	4.14.291	October 7, 2022	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockhol m), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfur t), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta) , Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)
2.0.2022 719.0	4.14.287	August 10, 2022	US West (N. California), Europe (Paris), Europe (Stockholm), Europe (Frankfurt), Asia Pacific (Mumbai), Middle East (Bahrain)

OsRelease Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20220426.0	4.14.281	June 14, 2022	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

Changes, Enhancements, and Resolved Issues

- Amazon EMR 5.36.0 upgrades now support: aws-sdk 1.12.206, Hadoop 2.10.1-amzn-4, Hive 2.3.9-amzn-2, Hudi 0.10.1-amzn-1, Spark 2.4.8-amzn-2, Presto 0.267-amzn-1, Amazon Glue connector 1.18.0, EMRFS 2.51.0.

Known issues

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&‘()*+,-`. For more information, see [UTF-8 encoding table and Unicode characters](https://www.utf8-chartable.de/).

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

- With Amazon EMR releases 5.36.0 and 6.6.0 through 6.9.0, `SecretAgent` and `RecordServer` service components may experience log data loss due to an incorrect file name pattern configuration in Log4j2 properties. The incorrect configuration causes the components to generate only one log file per day. When the rotation strategy occurs, it overwrites the existing file instead of generating a new log file as expected. As a workaround, use a bootstrap action to generate log files each hour and append an auto-increment integer in the file name to handle the rotation.

  For Amazon EMR 6.6.0 through 6.9.0 releases, use the following bootstrap action when you launch a cluster.

  ```
  --bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
  ```

  For Amazon EMR 5.36.0, use the following bootstrap action when you launch a cluster.

  ```
  --bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
  ```

5.36.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.16.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.5.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.21.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.7.0	EMR S3Select Connector

Component	Version	Description
emrfs	2.51.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.14.2	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.14.2	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.10.1-amzn-4	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.10.1-amzn-4	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.10.1-amzn-4	HDFS command-line client and library
hadoop-hdfs-namenode	2.10.1-amzn-4	HDFS service for tracking file names and block locations.

Component	Version	Description
hadoop-hdfs-journalnode	2.10.1-amzn-4	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.10.1-amzn-4	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.10.1-amzn-4	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.10.1-amzn-4	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.10.1-amzn-4	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.10.1-amzn-4	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.10.1-amzn-4	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.13	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.13	Service for serving one or more HBase regions.
hbase-client	1.4.13	HBase command-line client.

Component	Version	Description
hbase-rest-server	1.4.13	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.13	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.9-amzn-2	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.9-amzn-2	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.9-amzn-2	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.9-amzn-2	Hive command line client.
hive-hbase	2.3.9-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.9-amzn-2	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.9-amzn-2	Service for accepting Hive queries as web requests.
hudi	0.10.1-amzn-1	Incremental processing framework to power data pipeline at low latency and high efficiency.

Component	Version	Description
hudi-spark	0.10.1-amzn-1	Bundle library for running Spark with Hudi.
hudi-presto	0.10.1-amzn-1	Bundle library for running Presto with Hudi.
hue-server	4.10.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.4.1	Multi-user server for Jupyter notebooks
livy-server	0.7.1-incubating	REST interface for interacting with Apache Spark
nginx	1.13.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.8.0	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68	MySQL database server.
nvidia-cuda	11.0.194	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.5.0	Open Source Computer Vision Library.

Component	Version	Description
phoenix-library	4.14.3-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.3-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.267-amzn-1	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.267-amzn-1	Service for executing pieces of a query.
presto-client	0.267-amzn-1	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.8-amzn-2	Spark command-line clients.
spark-history-server	2.4.8-amzn-2	Web UI for viewing logged events for the lifetime of a completed Spark application.

Component	Version	Description
spark-on-yarn	2.4.8-amzn-2	In-memory execution engine for YARN.
spark-yarn-slave	2.4.8-amzn-2	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.4.1	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.10.0	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

5.36.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-5.36.0 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-ses sion.properties settings for Kubernetes/Yarn session.	Not available.
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Restarts Flink history server.

Classifications	Description	Reconfiguration Actions
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts PhoenixQu eryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.

Classifications	Description	Reconfiguration Actions
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	Should not be reconfigured.
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classifications	Description	Reconfiguration Actions
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.
hive-env	Change values in the Hive environment.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.

Classifications	Description	Reconfiguration Actions
hudi-env	Change values in the Hudi environment.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j	Change Livy log4j.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.

Classifications	Description	Reconfiguration Actions
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hb ase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServe r.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.pro perties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server.
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server.

Classifications	Description	Reconfiguration Actions
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server.
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server.
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-log4j	Change values in kms-log4j .properties file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
recordserver-env	Change values in the EMR RecordServer environment.	Restarts EMR record server.

Classifications	Description	Reconfiguration Actions
recordserver-conf	Change values in EMR RecordServer's erver.pro perties file.	Restarts EMR record server.
recordserver-log4j	Change values in EMR RecordServer's log4j.pro perties file.	Restarts EMR record server.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j	Change values in Spark's log4j.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restarts Oozie.

Classifications	Description	Reconfiguration Actions
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts MapReduce -HistoryServer.
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Livy Server and MapReduce-HistoryS erver.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

Amazon EMR release 5.35.0

5.35.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, Iceberg, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.35.0	**emr-5.34.0**	**emr-5.33.1**	**emr-5.33.0**
AWS SDK for Java	1.12.159	1.11.970	1.11.970	1.11.970
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.13.1	1.12.1	1.12.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.9	2.3.8	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.9	2.3.8	2.3.7	2.3.7
Hudi	0.9.0-amzn-2	0.9.0-amzn-0	0.7.0-amzn-1	0.7.0-amzn-1

	emr-5.35.0	emr-5.34.0	emr-5.33.1	emr-5.33.0
Hue	4.10.0	4.9.0	4.9.0	4.9.0
Iceberg	-	-	-	-
JupyterEn terpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.2.2	1.2.2
Livy	0.7.1	0.7.1	0.7.0	0.7.0
MXNet	1.8.0	1.8.0	1.7.0	1.7.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.1	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.266	0.261	0.245.1	0.245.1
Spark	2.4.8	2.4.8	2.4.7	2.4.7
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQ L)	-	-	-	-
Zeppelin	0.10.0	0.10.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

5.35.0 release notes

This is the Amazon EMR release 5.35.0 release note.

The following release notes include information for Amazon EMR release 5.35.0. Changes are relative to 5.34.0.

Initial release date: March 30, 2022

New Features

- Amazon EMR release 5.35 applications that use Log4j 1.x and Log4j 2.x are upgraded to use Log4j 1.2.17 (or higher) and Log4j 2.17.1 (or higher) respectively, and do not require using bootstrap actions to mitigate the CVE issues in previous releases. See Approach to mitigate CVE-2021-44228.

Changes, Enhancements, and Resolved Issues

Flink changes

Change type	Description
Upgrades	• Update flink version to 1.14.2. • log4j upgraded to 2.17.1.

Hadoop changes

Change type	Description
Hadoop open source backports since EMR 5.34.0	• YARN-10438: Handle null containerId in ClientRMService#getContainerReport() • YARN-7266: Timeline Server event handler threads locked • YARN-10438: ATS 1.5 fails to start if RollingLevelDb files are corrupt or missing • HADOOP-13500: Synchronizing iteration of Configuration properties object

Change type	Description
	• YARN-10651: CapacityScheduler crashed with NPE in AbstractYarnScheduler.updateNodeResource() • HDFS-12221: Replace xerces in XmlEditsVisitor • HDFS-16410: Insecure Xml parsing in OfflineEditsXmlLoader
Hadoop changes and fixes	• Tomcat used in KMS and HttpFS is upgraded to 8.5.75 • In FileSystemOptimizedCommitterV2, the success marker was written in the commitJob output path defined while creating the committer. Since commitJob and task level output paths can differ, the path has been corrected to use the one defined in manifest files. For Hive jobs, this results in the success marker being written correctly in when performing operations such as dynamic partition or UNION ALL.

Hive changes

Change type	Description
Hive upgraded to open source release 2.3.9, including these JIRA fixes	• HIVE-17155: findConfFile() in HiveConf.java has some issues with the conf path • HIVE-24797: Disable validate default values when parsing Avro schemas • HIVE-21563: Improve Table#getEmptyTable performance by disable registerAllFunctionsOnce • HIVE-18147: Tests can fail with java.net.BindException: Address already in use

Change type	Description
	• HIVE-24608: Switch back to get_table in HMS client for Hive 2.3.x • HIVE-21200: Vectorization - date column throwing java.lang.UnsupportedOperat ionException for parquet • HIVE-19228: Remove commons-httpclient 3.x usage
Hive open source backports since EMR 5.34.0	• HIVE-19990: Query with interval literal in join condition fails • HIVE-25824: Upgrade branch-2.3 to log4j 2.17.0 • TEZ-4062: Speculative attempt scheduling should be aborted when Task has completed • TEZ-4108: NullPointerException during speculative execution race condition • TEZ-3918: Setting tez.task.log.level does not work
Hive upgrades and fixes	• Upgrade Log4j version to 2.17.1 • Upgrade ORC version to 1.4.3 • FixED deadlock due to penalty thread in ShuffleScheduler
New features	• Added feature to print Hive Query in AM logs. This is disabled by default. Flag/Conf : `tez.am.emr.print.hive.query .in.log` . Status (default): FALSE.

Oozie changes

Change type	Description
Oozie open source backports since EMR 5.34.0	• OOZIE-3652: Oozie launcher should retry directory listing when NoSuchFileException occurs

Pig changes

Change type	Description
Upgrades	• log4j upgraded to 1.2.17.

Known issues

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&'()*+,-`. For more information, see UTF-8 encoding table and Unicode characters.

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

5.35.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.15.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.5.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.20.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.7.0	EMR S3Select Connector

Component	Version	Description
emrfs	2.49.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.14.2	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.14.2	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.10.1-amzn-3	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.10.1-amzn-3	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.10.1-amzn-3	HDFS command-line client and library
hadoop-hdfs-namenode	2.10.1-amzn-3	HDFS service for tracking file names and block locations.

Component	Version	Description
hadoop-hdfs-journalnode	2.10.1-amzn-3	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.10.1-amzn-3	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.10.1-amzn-3	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.10.1-amzn-3	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.10.1-amzn-3	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.10.1-amzn-3	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.10.1-amzn-3	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.13	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.13	Service for serving one or more HBase regions.
hbase-client	1.4.13	HBase command-line client.

Component	Version	Description
hbase-rest-server	1.4.13	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.13	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.9-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.9-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.9-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.9-amzn-0	Hive command line client.
hive-hbase	2.3.9-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.9-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.9-amzn-0	Service for accepting Hive queries as web requests.
hudi	0.9.0-amzn-2	Incremental processing framework to power data pipeline at low latency and high efficiency.

Component	Version	Description
hudi-spark	0.9.0-amzn-2	Bundle library for running Spark with Hudi.
hudi-presto	0.9.0-amzn-2	Bundle library for running Presto with Hudi.
hue-server	4.10.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.4.1	Multi-user server for Jupyter notebooks
livy-server	0.7.1-incubating	REST interface for interacting with Apache Spark
nginx	1.13.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.8.0	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68	MySQL database server.
nvidia-cuda	10.1.243	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.5.0	Open Source Computer Vision Library.

Component	Version	Description
phoenix-library	4.14.3-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.3-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.266-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.266-amzn-0	Service for executing pieces of a query.
presto-client	0.266-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.8-amzn-1	Spark command-line clients.
spark-history-server	2.4.8-amzn-1	Web UI for viewing logged events for the lifetime of a completed Spark application.

Component	Version	Description
spark-on-yarn	2.4.8-amzn-1	In-memory execution engine for YARN.
spark-yarn-slave	2.4.8-amzn-1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.4.1	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.10.0	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

5.35.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-5.35.0 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-ses sion.properties settings for Kubernetes/Yarn session.	Not available.
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Restarts Flink history server.

Classifications	Description	Reconfiguration Actions
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts PhoenixQu eryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.

Classifications	Description	Reconfiguration Actions
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	Should not be reconfigured.
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classifications	Description	Reconfiguration Actions
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.
hive-env	Change values in the Hive environment.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.

Classifications	Description	Reconfiguration Actions
hudi-env	Change values in the Hudi environment.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j	Change Livy log4j.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.

Classifications	Description	Reconfiguration Actions
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hb ase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServe r.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.pro perties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server.
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server.

Classifications	Description	Reconfiguration Actions
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server.
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server.
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-log4j	Change values in kms-log4j .properties file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
recordserver-env	Change values in the EMR RecordServer environment.	Restarts EMR record server.

Classifications	Description	Reconfiguration Actions
recordserver-conf	Change values in EMR RecordServer's erver.pro perties file.	Restarts EMR record server.
recordserver-log4j	Change values in EMR RecordServer's log4j.pro perties file.	Restarts EMR record server.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j	Change values in Spark's log4j.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restarts Oozie.

Classifications	Description	Reconfiguration Actions
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts MapReduce -HistoryServer.
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Livy Server and MapReduce-HistoryS erver.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

Amazon EMR release 5.34.0

5.34.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.34.0	emr-5.33.1	emr-5.33.0	emr-5.32.1
AWS SDK for Java	1.11.970	1.11.970	1.11.970	1.11.890
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.13.1	1.12.1	1.12.1	1.11.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.8	2.3.7	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.8	2.3.7	2.3.7	2.3.7
Hudi	0.9.0-amzn-0	0.7.0-amzn-1	0.7.0-amzn-1	0.6.0-amzn-0

	emr-5.34.0	emr-5.33.1	emr-5.33.0	emr-5.32.1
Hue	4.9.0	4.9.0	4.9.0	4.8.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.2.2	1.2.2	1.1.0
Livy	0.7.1	0.7.0	0.7.0	0.7.0
MXNet	1.8.0	1.7.0	1.7.0	1.7.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.261	0.245.1	0.245.1	0.240.1
Spark	2.4.8	2.4.7	2.4.7	2.4.7
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.3.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.10.0	0.9.0	0.9.0	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

5.34.0 release notes

The following release notes include information for Amazon EMR release 5.34.0. Changes are relative to 5.33.1.

Initial release date: January 20, 2022

Updated release date: March 21, 2022

New Features

- **[Managed scaling] Spark shuffle data managed scaling optimization** - For Amazon EMR versions 5.34.0 and later, and EMR versions 6.4.0 and later, managed scaling is now Spark shuffle data aware (data that Spark redistributes across partitions to perform specific operations). For more information on shuffle operations, see Using EMR managed scaling in Amazon EMR in the *Amazon EMR Management Guide* and Spark Programming Guide.
- [Hudi] Improvements to simplify Hudi configuration. Disabled optimistic concurrency control by default.

Changes, Enhancements, and Resolved Issues

- This is a release to fix issues with Amazon EMR Scaling when it fails to scale up/scale down a cluster successfully or causes application failures.
- Previously, manual restart of the resource manager on a multi-master cluster caused Amazon EMR on-cluster daemons, like Zookeeper, to reload all previously decommissioned or lost nodes in the Zookeeper znode file. This caused default limits to be exceeded in certain situations. Amazon EMR now removes the decommissioned or lost node records older than one hour from the Zookeeper file and the internal limits have been increased.
- Fixed an issue where scaling requests failed for a large, highly utilized cluster when Amazon EMR on-cluster daemons were running health checking activities, such as gathering YARN node state and HDFS node state. This was happening because on-cluster daemons were not able to communicate the health status data of a node to internal Amazon EMR components.
- Improved EMR on-cluster daemons to correctly track the node states when IP addresses are reused to improve reliability during scaling operations.
- SPARK-29683. Fixed an issue where job failures occurred during cluster scale-down as Spark was assuming all available nodes were deny-listed.

- YARN-9011. Fixed an issue where job failures occurred due to a race condition in YARN decommissioning when cluster tried to scale up or down.
- Fixed issue with step or job failures during cluster scaling by ensuring that the node states are always consistent between the Amazon EMR on-cluster daemons and YARN/HDFS.
- Fixed an issue where cluster operations such as scale down and step submission failed for Amazon EMR clusters enabled with Kerberos authentication. This was because the Amazon EMR on-cluster daemon did not renew the Kerberos ticket, which is required to securely communicate with HDFS/YARN running on the primary node.
- Zeppelin upgraded to version 0.10.0.
- Livy Fix - upgraded to 0.7.1
- Spark performance improvement - heterogeneous executors are disabled when certain Spark configuration values are overridden in EMR 5.34.0.
- WebHDFS and HttpFS server are disabled by default. You can re-enable WebHDFS using the Hadoop configuration, `dfs.webhdfs.enabled`. HttpFS server can be started by using `sudo systemctl start hadoop-httpfs`.

Known Issues

- The Amazon EMR Notebooks feature used with Livy user impersonation does not work because HttpFS is disabled by default. In this case, the EMR notebook cannot connect to the cluster that has Livy impersonation enabled. The workaround is to start HttpFS server before connecting the EMR notebook to the cluster using `sudo systemctl start hadoop-httpfs`.
- Hue queries do not work in Amazon EMR 6.4.0 because Apache Hadoop HttpFS server is disabled by default. To use Hue on Amazon EMR 6.4.0, either manually start HttpFS server on the Amazon EMR primary node using `sudo systemctl start hadoop-httpfs`, or use an Amazon EMR step.
- The Amazon EMR Notebooks feature used with Livy user impersonation does not work because HttpFS is disabled by default. In this case, the EMR notebook cannot connect to the cluster that has Livy impersonation enabled. The workaround is to start HttpFS server before connecting the EMR notebook to the cluster using `sudo systemctl start hadoop-httpfs`.
- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:

- Two or more partitions are scanned from the same table.
- At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
- The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&'()*+,-`. For more information, see UTF-8 encoding table and Unicode characters.

As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

5.34.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB connector for Hadoop ecosystem applications.

Component	Version	Description
emr-goodies	2.14.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.4.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.18.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.7.0	EMR S3Select Connector
emrfs	2.48.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.13.1	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.13.1	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.

Component	Version	Description
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.10.1-amzn-2	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.10.1-amzn-2	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.10.1-amzn-2	HDFS command-line client and library
hadoop-hdfs-namenode	2.10.1-amzn-2	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	2.10.1-amzn-2	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.10.1-amzn-2	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.10.1-amzn-2	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.10.1-amzn-2	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.10.1-amzn-2	YARN service for managing containers on an individual node.

Component	Version	Description
hadoop-yarn-resour cemanager	2.10.1-amzn-2	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.10.1-amzn-2	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.13	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.13	Service for serving one or more HBase regions.
hbase-client	1.4.13	HBase command-line client.
hbase-rest-server	1.4.13	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.13	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.8-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.8-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.8-amzn-0	HTTP endpoint providing a REST interface to HCatalog.

Component	Version	Description
hive-client	2.3.8-amzn-0	Hive command line client.
hive-hbase	2.3.8-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.8-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.8-amzn-0	Service for accepting Hive queries as web requests.
hudi	0.9.0-amzn-0	Incremental processing framework to power data pipeline at low latency and high efficiency.
hudi-spark	0.9.0-amzn-0	Bundle library for running Spark with Hudi.
hudi-presto	0.9.0-amzn-0	Bundle library for running Presto with Hudi.
hue-server	4.9.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.4.1	Multi-user server for Jupyter notebooks
livy-server	0.7.1-incubating	REST interface for interacting with Apache Spark
nginx	1.13.1	nginx [engine x] is an HTTP and reverse proxy server

Component	Version	Description
mahout-client	0.13.0	Library for machine learning.
mxnet	1.8.0	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68	MySQL database server.
nvidia-cuda	10.1.243	Nvidia drivers and Cuda toolkit
oozie-client	5.2.1	Oozie command-line client.
oozie-server	5.2.1	Service for accepting Oozie workflow requests.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.3-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.261-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.261-amzn-0	Service for executing pieces of a query.

Component	Version	Description
presto-client	0.261-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.8-amzn-0	Spark command-line clients.
spark-history-server	2.4.8-amzn-0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.8-amzn-0	In-memory execution engine for YARN.
spark-yarn-slave	2.4.8-amzn-0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.4.1	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.

Component	Version	Description
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.10.0	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

5.34.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-5.34.0 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.

Classifications	Description	Reconfiguration Actions
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.

Classifications	Description	Reconfiguration Actions
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-session.properties settings for Kubernetes/Yarn session.	Not available.
flink-log4j-cli	Change Flink log4j-cli.properties settings.	Restarts Flink history server.

Classifications	Description	Reconfiguration Actions
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts PhoenixQu eryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.

Classifications	Description	Reconfiguration Actions
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	Should not be reconfigured.
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classifications	Description	Reconfiguration Actions
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.
hive-env	Change values in the Hive environment.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.

Classifications	Description	Reconfiguration Actions
hudi-env	Change values in the Hudi environment.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j	Change Livy log4j.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.

Classifications	Description	Reconfiguration Actions
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServer.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server.
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server.

Classifications	Description	Reconfiguration Actions
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server.
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server.
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-log4j	Change values in kms-log4j .properties file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
recordserver-env	Change values in the EMR RecordServer environment.	Restarts EMR record server.

Classifications	Description	Reconfiguration Actions
recordserver-conf	Change values in EMR RecordServer's erver.pro perties file.	Restarts EMR record server.
recordserver-log4j	Change values in EMR RecordServer's log4j.pro perties file.	Restarts EMR record server.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j	Change values in Spark's log4j.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restarts Oozie.

Classifications	Description	Reconfiguration Actions
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts MapReduce -HistoryServer.
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Livy Server and MapReduce-HistoryS erver.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

Amazon EMR release 5.33.1

5.33.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.33.1	emr-5.33.0	emr-5.32.1	emr-5.32.0
AWS SDK for Java	1.11.970	1.11.970	1.11.890	1.11.890
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.12.1	1.12.1	1.11.2	1.11.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.7	2.3.7	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.7	2.3.7	2.3.7	2.3.7
Hudi	0.7.0-amzn-1	0.7.0-amzn-1	0.6.0-amzn-0	0.6.0-amzn-0

	emr-5.33.1	emr-5.33.0	emr-5.32.1	emr-5.32.0
Hue	4.9.0	4.9.0	4.8.0	4.8.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.2.2	1.2.2	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.7.0	1.7.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.245.1	0.245.1	0.240.1	0.240.1
Spark	2.4.7	2.4.7	2.4.7	2.4.7
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.3.1	2.3.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.9.0	0.9.0	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

5.33.1 release notes

The following release notes include information for Amazon EMR release 5.33.0/5.33.1. Changes are relative to 5.32.0.

Initial release date: April 19, 2021

Last updated date: August 9, 2021

Upgrades

- Upgraded Amazon Glue connector to version 1.15.0
- Upgraded AWS SDK for Java to version 1.11.970
- Upgraded EMRFS to version 2.46.0
- Upgraded EMR Goodies to version 2.14.0
- Upgraded EMR Record Server to version 1.9.0
- Upgraded EMR S3 Dist CP to version 2.18.0
- Upgraded EMR Secret Agent to version 1.8.0
- Upgraded Flink to version 1.12.1
- Upgraded Hadoop to version 2.10.1-amzn-1
- Upgraded Hive to version 2.3.7-amzn-4
- Upgraded Hudi to version 0.7.0
- Upgraded Hue to version 4.9.0
- Upgraded OpenCV to version 4.5.0
- Upgraded Presto to version 0.245.1-amzn-0
- Upgraded R to version 4.0.2
- Upgraded Spark to version 2.4.7-amzn-1
- Upgraded TensorFlow to version 2.4.1
- Upgraded Zeppelin to version 0.9.0

Changes, enhancements, and resolved issues

- This is a release to fix issues with Amazon EMR Scaling when it fails to scale up/scale down a cluster successfully or causes application failures.

- Fixed an issue where scaling requests failed for a large, highly utilized cluster when Amazon EMR on-cluster daemons were running health checking activities, such as gathering YARN node state and HDFS node state. This was happening because on-cluster daemons were not able to communicate the health status data of a node to internal Amazon EMR components.
- Improved EMR on-cluster daemons to correctly track the node states when IP addresses are reused to improve reliability during scaling operations.
- SPARK-29683. Fixed an issue where job failures occurred during cluster scale-down as Spark was assuming all available nodes were deny-listed.
- YARN-9011. Fixed an issue where job failures occurred due to a race condition in YARN decommissioning when cluster tried to scale up or down.
- Fixed issue with step or job failures during cluster scaling by ensuring that the node states are always consistent between the Amazon EMR on-cluster daemons and YARN/HDFS.
- Fixed an issue where cluster operations such as scale down and step submission failed for Amazon EMR clusters enabled with Kerberos authentication. This was because the Amazon EMR on-cluster daemon did not renew the Kerberos ticket, which is required to securely communicate with HDFS/YARN running on the primary node.
- Newer Amazon EMR releases fix the issue with a lower "Max open files" limit on older AL2 in Amazon EMR. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later now include a permanent fix with a higher "Max open files" setting.
- **Configuring a cluster to fix Apache YARN Timeline Server version 1 and 1.5 performance issues**

  Apache YARN Timeline Server version 1 and 1.5 can cause performance issues with very active, large EMR clusters, particularly with `yarn.resourcemanager.system-metrics-publisher.enabled=true`, which is the default setting in Amazon EMR. An open source YARN Timeline Server v2 solves the performance issue related to YARN Timeline Server scalability.

  Other workarounds for this issue include:

  - Configuring yarn.resourcemanager.system-metrics-publisher.enabled=false in yarn-site.xml.
  - Enabling the fix for this issue when creating a cluster, as described below.

  The following Amazon EMR releases contain a fix for this YARN Timeline Server performance issue.

  EMR 5.30.2, 5.31.1, 5.32.1, 5.33.1, 5.34.x, 6.0.1, 6.1.1, 6.2.1, 6.3.1, 6.4.x

To enable the fix on any of the above specified Amazon EMR releases, set these properties to `true` in a configurations JSON file that is passed in using the `aws emr create-cluster` command parameter: `--configurations file://./configurations.json`. Or enable the fix using the reconfiguration console UI.

Example of the configurations.json file contents:

```
[
{
"Classification": "yarn-site",
"Properties": {
"yarn.resourcemanager.system-metrics-publisher.timeline-server-v1.enable-batch":
 "true",
"yarn.resourcemanager.system-metrics-publisher.enabled": "true"
},
"Configurations": []
}
]
```

- Spark runtime is now faster when fetching partition locations from Hive Metastore for Spark insert queries.
- Upgraded component versions. For a list of component versions, see About Amazon EMR Releases in this guide.
- Installed the AWS Java SDK Bundle on each new cluster. This is a single jar containing all service SDKs and their dependencies, instead of individual component jars. For more information, see Java SDK Bundled Dependency.
- Fixed Managed Scaling issues in earlier Amazon EMR releases and made improvements so application failure rates are significantly reduced.
- HTTPS is now enabled by default for Amazon Linux repositories. If you are using an Amazon S3 VPCE policy to restrict access to specific buckets, you must add the new Amazon Linux bucket ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` to your policy (replace `$region` with the region where the endpoint is). For more information, see this topic in the AWS discussion forums. Announcement: Amazon Linux 2 now supports the ability to use HTTPS while connecting to package repositories .

New features

- Amazon EMR supports Amazon S3 Access Points, a feature of Amazon S3 that allows you to easily manage access for shared data lakes. Using your Amazon S3 Access Point alias, you can simplify your data access at scale on Amazon EMR. You can use Amazon S3 Access Points with all versions of Amazon EMR at no additional cost in all AWS regions where Amazon EMR is available. To learn more about Amazon S3 Access Points and Access Point aliases, see Using a bucket-style alias for your access point in the *Amazon S3 User Guide*.
- Amazon EMR-5.33 supports new Amazon EC2 instance types: c5a, c5ad, c6gn, c6gd, m6gd, d3, d3en, m5zn, r5b, r6gd. See Supported Instance Types.

Known issues

- **Lower "Max open files" limit on older AL2 [fixed in newer releases].** Amazon EMR releases: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0, and emr-6.2.0 are based on older versions ofAmazon Linux 2 (AL2), which have a lower ulimit setting for "Max open files" when Amazon EMR clusters are created with the default AMI. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later include a permanent fix with a higher "Max open files" setting. Releases with the lower open file limit causes a "Too many open files" error when submitting Spark job. In the impacted releases, the Amazon EMR default AMI has a default ulimit setting of 4096 for "Max open files," which is lower than the 65536 file limit in the latestAmazon Linux 2 AMI. The lower ulimit setting for "Max open files" causes Spark job failure when the Spark driver and executor try to open more than 4096 files. To fix the issue, Amazon EMR has a bootstrap action (BA) script that adjusts the ulimit setting at cluster creation.

  If you are using an older Amazon EMR version that doesn't have the permanent fix for this issue, the following workaround lets you to explicitly set the instance-controller ulimit to a maximum of 65536 files.

  Explicitly set a ulimit from the command line

  1. Edit `/etc/systemd/system/instance-controller.service` to add the following parameters to Service section.

     `LimitNOFILE=65536`

     `LimitNPROC=65536`
  2. Restart InstanceController

```
$ sudo systemctl daemon-reload

$ sudo systemctl restart instance-controller
```

Set a ulimit using bootstrap action (BA)

You can also use a bootstrap action (BA) script to configure the instance-controller ulimit to 65536 files at cluster creation.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- For Amazon EMR 6.3.0 and 6.2.0 private subnet clusters, you cannot access the Ganglia web UI. You will get an "access denied (403)" error. Other web UIs, such as Spark, Hue, JupyterHub, Zeppelin, Livy, and Tez are working normally. Ganglia web UI access on public subnet clusters are also working normally. To resolve this issue, restart httpd service on the primary node with `sudo systemctl restart httpd`. This issue is fixed in Amazon EMR 6.4.0.
- **Important**

  EMR clusters that run Amazon Linux or Amazon Linux 2 Amazon Machine Images (AMIs) use default Amazon Linux behavior, and do not automatically download and install important and critical kernel updates that require a reboot. This is the same behavior as other Amazon EC2 instances that run the default Amazon Linux AMI. If new Amazon

Linux software updates that require a reboot (such as kernel, NVIDIA, and CUDA updates) become available after an Amazon EMR release becomes available, EMR cluster instances that run the default AMI do not automatically download and install those updates. To get kernel updates, you can customize your Amazon EMR AMI to use the latest Amazon Linux AMI.

- Console support to create a security configuration that specifies the AWS Ranger integration option is currently not supported in the GovCloud Region. Security configuration can be done using the CLI. See Create the EMR Security Configuration in the *Amazon EMR Management Guide*.
- Scoped managed policies: To align with AWS best practices, Amazon EMR has introduced v2 EMR-scoped default managed policies as replacements for policies that will be deprecated. See Amazon EMR Managed Policies.
- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&‘()*+,-`. For more information, see UTF-8 encoding table and Unicode characters.

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

5.33.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.14.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.2.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.18.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.6.0	EMR S3Select Connector
emrfs	2.46.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.12.1	Apache Flink command line client scripts and applications.

Component	Version	Description
flink-jobmanager-config	1.12.1	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.10.1-amzn-1.1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.10.1-amzn-1.1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.10.1-amzn-1.1	HDFS command-line client and library
hadoop-hdfs-namenode	2.10.1-amzn-1.1	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	2.10.1-amzn-1.1	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.10.1-amzn-1.1	HTTP endpoint for HDFS operations.

Component	Version	Description
hadoop-kms-server	2.10.1-amzn-1.1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.10.1-amzn-1.1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.10.1-amzn-1.1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.10.1-amzn-1.1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.10.1-amzn-1.1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.13	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.13	Service for serving one or more HBase regions.
hbase-client	1.4.13	HBase command-line client.
hbase-rest-server	1.4.13	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.13	Service providing a Thrift endpoint to HBase.

Component	Version	Description
hcatalog-client	2.3.7-amzn-4	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.7-amzn-4	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.7-amzn-4	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.7-amzn-4	Hive command line client.
hive-hbase	2.3.7-amzn-4	Hive-hbase client.
hive-metastore-server	2.3.7-amzn-4	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.7-amzn-4	Service for accepting Hive queries as web requests.
hudi	0.7.0-amzn-1	Incremental processing framework to power data pipline at low latency and high efficiency.
hudi-spark	0.7.0-amzn-1	Bundle library for running Spark with Hudi.
hudi-presto	0.7.0-amzn-1	Bundle library for running Presto with Hudi.

Component	Version	Description
hue-server	4.9.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.2.2	Multi-user server for Jupyter notebooks
livy-server	0.7.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.7.0	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68+	MySQL database server.
nvidia-cuda	10.1.243	Nvidia drivers and Cuda toolkit
oozie-client	5.2.0	Oozie command-line client.
oozie-server	5.2.0	Service for accepting Oozie workflow requests.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	The phoenix libraries for server and client

Component	Version	Description
phoenix-query-server	4.14.3-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.245.1-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.245.1-amzn-0	Service for executing pieces of a query.
presto-client	0.245.1-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.7-amzn-1.1	Spark command-line clients.
spark-history-server	2.4.7-amzn-1.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.7-amzn-1.1	In-memory execution engine for YARN.

Component	Version	Description
spark-yarn-slave	2.4.7-amzn-1.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.4.1	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.9.0	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

5.33.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-5.33.1 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode,

Classifications	Description	Reconfiguration Actions
		SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-ses sion.properties settings for Kubernetes/Yarn session.	Not available.
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Restarts Flink history server.

Classifications	Description	Reconfiguration Actions
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts PhoenixQu eryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.

Classifications	Description	Reconfiguration Actions
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	Should not be reconfigured.
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classifications	Description	Reconfiguration Actions
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.
hive-env	Change values in the Hive environment.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.

Classifications	Description	Reconfiguration Actions
hudi-env	Change values in the Hudi environment.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j	Change Livy log4j.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.

Classifications	Description	Reconfiguration Actions
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hb ase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServe r.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.pro perties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server.
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server.

Classifications	Description	Reconfiguration Actions
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server.
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server.
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-log4j	Change values in kms-log4j .properties file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
recordserver-env	Change values in the EMR RecordServer environment.	Restarts EMR record server.

Classifications	Description	Reconfiguration Actions
recordserver-conf	Change values in EMR RecordServer's erver.pro perties file.	Restarts EMR record server.
recordserver-log4j	Change values in EMR RecordServer's log4j.pro perties file.	Restarts EMR record server.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j	Change values in Spark's log4j.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restarts Oozie.

Classifications	Description	Reconfiguration Actions
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts MapReduce -HistoryServer.
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Livy Server and MapReduce-HistoryS erver.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

Amazon EMR release 5.33.0

5.33.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.33.0	emr-5.32.1	emr-5.32.0	emr-5.31.1
AWS SDK for Java	1.11.970	1.11.890	1.11.890	1.11.852
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.12.1	1.11.2	1.11.2	1.11.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.7	2.3.7	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.1	2.10.0
Hive	2.3.7	2.3.7	2.3.7	2.3.7
Hudi	0.7.0-amzn-1	0.6.0-amzn-0	0.6.0-amzn-0	0.6.0-amzn-0

	emr-5.33.0	emr-5.32.1	emr-5.32.0	emr-5.31.1
Hue	4.9.0	4.8.0	4.8.0	4.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	-
JupyterHub	1.2.2	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.7.0	1.6.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.245.1	0.240.1	0.240.1	0.238.3
Spark	2.4.7	2.4.7	2.4.7	2.4.6
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.3.1	2.3.1	2.1.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.9.0	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

5.33.0 release notes

5.33.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.14.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.2.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.18.0	Distributed copy application optimized for Amazon S3.

Component	Version	Description
emr-s3-select	1.6.0	EMR S3Select Connector
emrfs	2.46.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.12.1	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.12.1	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.10.1-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.10.1-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.10.1-amzn-1	HDFS command-line client and library

Component	Version	Description
hadoop-hdfs-namenode	2.10.1-amzn-1	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	2.10.1-amzn-1	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.10.1-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.10.1-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.10.1-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.10.1-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.10.1-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.10.1-amzn-1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.13	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.

Component	Version	Description
hbase-region-server	1.4.13	Service for serving one or more HBase regions.
hbase-client	1.4.13	HBase command-line client.
hbase-rest-server	1.4.13	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.13	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.7-amzn-4	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.7-amzn-4	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.7-amzn-4	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.7-amzn-4	Hive command line client.
hive-hbase	2.3.7-amzn-4	Hive-hbase client.
hive-metastore-server	2.3.7-amzn-4	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.7-amzn-4	Service for accepting Hive queries as web requests.

Component	Version	Description
hudi	0.7.0-amzn-1	Incremental processing framework to power data pipline at low latency and high efficiency.
hudi-spark	0.7.0-amzn-1	Bundle library for running Spark with Hudi.
hudi-presto	0.7.0-amzn-1	Bundle library for running Presto with Hudi.
hue-server	4.9.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.2.2	Multi-user server for Jupyter notebooks
livy-server	0.7.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.7.0	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68	MySQL database server.
nvidia-cuda	10.1.243	Nvidia drivers and Cuda toolkit
oozie-client	5.2.0	Oozie command-line client.

Component	Version	Description
oozie-server	5.2.0	Service for accepting Oozie workflow requests.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.3-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.245.1-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.245.1-amzn-0	Service for executing pieces of a query.
presto-client	0.245.1-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
pig-client	0.17.0	Pig command-line client.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Apache Ranger Key Management System

Component	Version	Description
spark-client	2.4.7-amzn-1	Spark command-line clients.
spark-history-server	2.4.7-amzn-1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.7-amzn-1	In-memory execution engine for YARN.
spark-yarn-slave	2.4.7-amzn-1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.4.1	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.9.0	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.

Component	Version	Description
zookeeper-client	3.4.14	ZooKeeper command line client.

5.33.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-5.33.0 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad

Classifications	Description	Reconfiguration Actions
		ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-session	Change Flink log4j-ses sion.properties settings for Kubernetes/Yarn session.	Not available.

Classifications	Description	Reconfiguration Actions
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Restarts Flink history server.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.

Classifications	Description	Reconfiguration Actions
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	Should not be reconfigured.
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classifications	Description	Reconfiguration Actions
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.
hive-env	Change values in the Hive environment.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.

Classifications	Description	Reconfiguration Actions
hudi-env	Change values in the Hudi environment.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j	Change Livy log4j.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.

Classifications	Description	Reconfiguration Actions
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServer.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server.
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server.

Classifications	Description	Reconfiguration Actions
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server.
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server.
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-log4j	Change values in kms-log4j .properties file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
recordserver-env	Change values in the EMR RecordServer environment.	Restarts EMR record server.

Classifications	Description	Reconfiguration Actions
recordserver-conf	Change values in EMR RecordServer's erver.pro perties file.	Restarts EMR record server.
recordserver-log4j	Change values in EMR RecordServer's log4j.pro perties file.	Restarts EMR record server.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j	Change values in Spark's log4j.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restarts Oozie.

Classifications	Description	Reconfiguration Actions
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts MapReduce -HistoryServer.
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Livy Server and MapReduce-HistoryS erver.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zeppelin-site	Change configuration settings in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

Amazon EMR release 5.32.1

5.32.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.32.1	emr-5.32.0	emr-5.31.1	emr-5.31.0
AWS SDK for Java	1.11.890	1.11.890	1.11.852	1.11.852
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.2	1.11.2	1.11.0	1.11.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.7	2.3.7	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.0	2.10.0
Hive	2.3.7	2.3.7	2.3.7	2.3.7
Hudi	0.6.0-amzn-0	0.6.0-amzn-0	0.6.0-amzn-0	0.6.0-amzn-0

	emr-5.32.1	emr-5.32.0	emr-5.31.1	emr-5.31.0
Hue	4.8.0	4.8.0	4.7.1	4.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.6.0	1.6.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.240.1	0.240.1	0.238.3	0.238.3
Spark	2.4.7	2.4.7	2.4.6	2.4.6
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.3.1	2.3.1	2.1.0	2.1.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

5.32.1 release notes

This is a release to fix issues with Amazon EMR Scaling when it fails to scale up/scale down a cluster successfully or causes application failures.

Changes, Enhancements, and Resolved Issues

- Fixed an issue where scaling requests failed for a large, highly utilized cluster when Amazon EMR on-cluster daemons were running health checking activities, such as gathering YARN node state and HDFS node state. This was happening because on-cluster daemons were not able to communicate the health status data of a node to internal Amazon EMR components.
- Improved EMR on-cluster daemons to correctly track the node states when IP addresses are reused to improve reliability during scaling operations.
- SPARK-29683. Fixed an issue where job failures occurred during cluster scale-down as Spark was assuming all available nodes were deny-listed.
- YARN-9011. Fixed an issue where job failures occurred due to a race condition in YARN decommissioning when cluster tried to scale up or down.
- Fixed issue with step or job failures during cluster scaling by ensuring that the node states are always consistent between the Amazon EMR on-cluster daemons and YARN/HDFS.
- Fixed an issue where cluster operations such as scale down and step submission failed for Amazon EMR clusters enabled with Kerberos authentication. This was because the Amazon EMR on-cluster daemon did not renew the Kerberos ticket, which is required to securely communicate with HDFS/YARN running on the primary node.
- Newer Amazon EMR releases fix the issue with a lower "Max open files" limit on older AL2 in Amazon EMR. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later now include a permanent fix with a higher "Max open files" setting.
- HTTPS is now enabled by default for Amazon Linux repositories. If you are using an Amazon S3 VPCE policy to restrict access to specific buckets, you must add the new Amazon Linux bucket ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` to your policy (replace `$region` with the region where the endpoint is). For more information, see this topic in the AWS discussion forums. Announcement: Amazon Linux 2 now supports the ability to use HTTPS while connecting to package repositories .

Known issues

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&‘()*+,-`. For more information, see UTF-8 encoding table and Unicode characters.

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

5.32.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK

Component	Version	Description
emr-ddb	4.16.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.13.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.1.0	Conda env for emr notebook which includes jupyter enterprise gateway
emr-s3-dist-cp	2.17.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.6.0	EMR S3Select Connector
emrfs	2.45.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.11.2	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.11.2	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.

Component	Version	Description
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.10.1-amzn-0.1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.10.1-amzn-0.1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.10.1-amzn-0.1	HDFS command-line client and library
hadoop-hdfs-namenode	2.10.1-amzn-0.1	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	2.10.1-amzn-0.1	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.10.1-amzn-0.1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.10.1-amzn-0.1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.10.1-amzn-0.1	MapReduce execution engine libraries for running a MapReduce application.

Component	Version	Description
hadoop-yarn-nodemanager	2.10.1-amzn-0.1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.10.1-amzn-0.1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.10.1-amzn-0.1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.13	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.13	Service for serving one or more HBase regions.
hbase-client	1.4.13	HBase command-line client.
hbase-rest-server	1.4.13	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.13	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.7-amzn-3	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.7-amzn-3	Service providing HCatalog, a table and storage management layer for distributed applications.

Component	Version	Description
hcatalog-webhcat-server	2.3.7-amzn-3	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.7-amzn-3	Hive command line client.
hive-hbase	2.3.7-amzn-3	Hive-hbase client.
hive-metastore-server	2.3.7-amzn-3	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.7-amzn-3	Service for accepting Hive queries as web requests.
hudi	0.6.0-amzn-0	Incremental processing framework to power data pipline at low latency and high efficiency.
hudi-spark	0.6.0-amzn-0	Bundle library for running Spark with Hudi.
hudi-presto	0.6.0-amzn-0	Bundle library for running Presto with Hudi.
hue-server	4.8.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.1.0	Multi-user server for Jupyter notebooks
livy-server	0.7.0-incubating	REST interface for interacting with Apache Spark

Component	Version	Description
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.7.0	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68+	MySQL database server.
nvidia-cuda	10.1.243	Nvidia drivers and Cuda toolkit
oozie-client	5.2.0	Oozie command-line client.
oozie-server	5.2.0	Service for accepting Oozie workflow requests.
opencv	4.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.3-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.240.1-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.

Component	Version	Description
presto-worker	0.240.1-amzn-0	Service for executing pieces of a query.
presto-client	0.240.1-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
pig-client	0.17.0	Pig command-line client.
r	3.4.3	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.7-amzn-0.1	Spark command-line clients.
spark-history-server	2.4.7-amzn-0.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.7-amzn-0.1	In-memory execution engine for YARN.
spark-yarn-slave	2.4.7-amzn-0.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.3.1	TensorFlow open source software library for high performance numerical computation.

Component	Version	Description
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.2	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

5.32.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-5.32.1 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.

Classifications	Description	Reconfiguration Actions
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.

Classifications	Description	Reconfiguration Actions
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.	Not available.
flink-log4j-cli	Change Flink log4j-cli.properties settings.	Restarts Flink history server.

Classifications	Description	Reconfiguration Actions
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts PhoenixQu eryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.

Classifications	Description	Reconfiguration Actions
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	Should not be reconfigured.
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classifications	Description	Reconfiguration Actions
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.
hive-env	Change values in the Hive environment.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.

Classifications	Description	Reconfiguration Actions
hudi-env	Change values in the Hudi environment.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j	Change Livy log4j.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.

Classifications	Description	Reconfiguration Actions
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hb ase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServe r.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.pro perties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server.
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server.

Classifications	Description	Reconfiguration Actions
presto-password-authenticat or	Change values in Presto's password-authenticator.prop erties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server.
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server.
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-log4j	Change values in kms-log4j .properties file of Ranger KMS.	Not available.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
recordserver-env	Change values in the EMR RecordServer environment.	Restarts EMR record server.

Classifications	Description	Reconfiguration Actions
recordserver-conf	Change values in EMR RecordServer's erver.pro perties file.	Restarts EMR record server.
recordserver-log4j	Change values in EMR RecordServer's log4j.pro perties file.	Restarts EMR record server.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j	Change values in Spark's log4j.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restarts Oozie.

Classifications	Description	Reconfiguration Actions
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts MapReduce -HistoryServer.
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Livy Server and MapReduce-HistoryS erver.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

Amazon EMR release 5.32.0

5.32.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, JupyterEnterpriseGateway, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.32.0	emr-5.31.1	emr-5.31.0	emr-5.30.2
AWS SDK for Java	1.11.890	1.11.852	1.11.852	1.11.759
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.2	1.11.0	1.11.0	1.10.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.7	2.3.7	2.3.7	2.3.6
Hadoop	2.10.1	2.10.0	2.10.0	2.8.5
Hive	2.3.7	2.3.7	2.3.7	2.3.6
Hudi	0.6.0-amzn-0	0.6.0-amzn-0	0.6.0-amzn-0	0.5.2-incubating
Hue	4.8.0	4.7.1	4.7.1	4.6.0

	emr-5.32.0	emr-5.31.1	emr-5.31.0	emr-5.30.2
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.6.0	1.6.0	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.240.1	0.238.3	0.238.3	0.232
Spark	2.4.7	2.4.6	2.4.6	2.4.5
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.3.1	2.1.0	2.1.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

5.32.0 release notes

The following release notes include information for Amazon EMR release 5.32.0. Changes are relative to 5.31.0.

Initial release date: Jan 8, 2021

Upgrades

- Upgraded Amazon Glue connector to version 1.14.0
- Upgraded Amazon SageMaker Spark SDK to version 1.4.1
- Upgraded AWS SDK for Java to version 1.11.890
- Upgraded EMR DynamoDB Connector version 4.16.0
- Upgraded EMRFS to version 2.45.0
- Upgraded EMR Log Analytics Metrics to version 1.18.0
- Upgraded EMR MetricsAndEventsApiGateway Client to version 1.5.0
- Upgraded EMR Record Server to version 1.8.0
- Upgraded EMR S3 Dist CP to version 2.17.0
- Upgraded EMR Secret Agent to version 1.7.0
- Upgraded Flink to version 1.11.2
- Upgraded Hadoop to version 2.10.1-amzn-0
- Upgraded Hive to version 2.3.7-amzn-3
- Upgraded Hue to version 4.8.0
- Upgraded Mxnet to version 1.7.0
- Upgraded OpenCV to version 4.4.0
- Upgraded Presto to version 0.240.1-amzn-0
- Upgraded Spark to version 2.4.7-amzn-0
- Upgraded TensorFlow to version 2.3.1

Changes, enhancements, and resolved issues

- This is a release to fix issues with Amazon EMR Scaling when it fails to scale up/scale down a cluster successfully or causes application failures.

- Fixed an issue where scaling requests failed for a large, highly utilized cluster when Amazon EMR on-cluster daemons were running health checking activities, such as gathering YARN node state and HDFS node state. This was happening because on-cluster daemons were not able to communicate the health status data of a node to internal Amazon EMR components.
- Improved EMR on-cluster daemons to correctly track the node states when IP addresses are reused to improve reliability during scaling operations.
- SPARK-29683. Fixed an issue where job failures occurred during cluster scale-down as Spark was assuming all available nodes were deny-listed.
- YARN-9011. Fixed an issue where job failures occurred due to a race condition in YARN decommissioning when cluster tried to scale up or down.
- Fixed issue with step or job failures during cluster scaling by ensuring that the node states are always consistent between the Amazon EMR on-cluster daemons and YARN/HDFS.
- Fixed an issue where cluster operations such as scale down and step submission failed for Amazon EMR clusters enabled with Kerberos authentication. This was because the Amazon EMR on-cluster daemon did not renew the Kerberos ticket, which is required to securely communicate with HDFS/YARN running on the primary node.
- Newer Amazon EMR releases fix the issue with a lower "Max open files" limit on older AL2 in Amazon EMR. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later now include a permanent fix with a higher "Max open files" setting.
- Upgraded component versions.
- For a list of component versions, see About Amazon EMR Releases in this guide.

New features

- Starting with Amazon EMR 5.32.0 and 6.5.0, dynamic executor sizing for Apache Spark is enabled by default. To turn this feature on or off, you can use the `spark.yarn.heterogeneousExecutors.enabled` configuration parameter.
- Instance Metadata Service (IMDS) V2 support status: Amazon EMR 5.23.1, 5.27.1 and 5.32 or later components use IMDSv2 for all IMDS calls. For IMDS calls in your application code, you can use both IMDSv1 and IMDSv2, or configure the IMDS to use only IMDSv2 for added security. For other 5.x EMR releases, disabling IMDSv1 causes cluster startup failure.
- Beginning with Amazon EMR 5.32.0, you can launch a cluster that natively integrates with Apache Ranger. Apache Ranger is an open-source framework to enable, monitor, and manage comprehensive data security across the Hadoop platform. For more information, see Apache

Ranger. With native integration, you can bring your own Apache Ranger to enforce fine-grained data access control on Amazon EMR. See Integrate Amazon EMR with Apache Ranger in the *Amazon EMR Release Guide*.

- Amazon EMR Release 5.32.0 supports Amazon EMR on EKS. For more details on getting started with EMR on EKS, see What is Amazon EMR on EKS.
- Amazon EMR Release 5.32.0 supports Amazon EMR Studio (Preview). For more details on getting started with EMR Studio, see Amazon EMR Studio (Preview).
- Scoped managed policies: To align with AWS best practices, Amazon EMR has introduced v2 EMR-scoped default managed policies as replacements for policies that will be deprecated. See Amazon EMR Managed Policies.

Known issues

- For Amazon EMR 6.3.0 and 6.2.0 private subnet clusters, you cannot access the Ganglia web UI. You will get an "access denied (403)" error. Other web UIs, such as Spark, Hue, JupyterHub, Zeppelin, Livy, and Tez are working normally. Ganglia web UI access on public subnet clusters are also working normally. To resolve this issue, restart httpd service on the primary node with `sudo systemctl restart httpd`. This issue is fixed in Amazon EMR 6.4.0.
- **Lower "Max open files" limit on older AL2 [fixed in newer releases].** Amazon EMR releases: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0, and emr-6.2.0 are based on older versions ofAmazon Linux 2 (AL2), which have a lower ulimit setting for "Max open files" when Amazon EMR clusters are created with the default AMI. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later include a permanent fix with a higher "Max open files" setting. Releases with the lower open file limit causes a "Too many open files" error when submitting Spark job. In the impacted releases, the Amazon EMR default AMI has a default ulimit setting of 4096 for "Max open files," which is lower than the 65536 file limit in the latestAmazon Linux 2 AMI. The lower ulimit setting for "Max open files" causes Spark job failure when the Spark driver and executor try to open more than 4096 files. To fix the issue, Amazon EMR has a bootstrap action (BA) script that adjusts the ulimit setting at cluster creation.

  If you are using an older Amazon EMR version that doesn't have the permanent fix for this issue, the following workaround lets you to explicitly set the instance-controller ulimit to a maximum of 65536 files.

Explicitly set a ulimit from the command line

1. Edit `/etc/systemd/system/instance-controller.service` to add the following parameters to Service section.

   `LimitNOFILE=65536`

   `LimitNPROC=65536`

2. Restart InstanceController

   `$ sudo systemctl daemon-reload`

   `$ sudo systemctl restart instance-controller`

Set a ulimit using bootstrap action (BA)

You can also use a bootstrap action (BA) script to configure the instance-controller ulimit to 65536 files at cluster creation.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- > **Important**
  >
  > EMR clusters that run Amazon Linux or Amazon Linux 2 Amazon Machine Images (AMIs) use default Amazon Linux behavior, and do not automatically download and install important and critical kernel updates that require a reboot. This is the same behavior as other Amazon EC2 instances that run the default Amazon Linux AMI. If new Amazon Linux software updates that require a reboot (such as kernel, NVIDIA, and CUDA updates) become available after an Amazon EMR release becomes available, EMR cluster instances that run the default AMI do not automatically download and install those updates. To get kernel updates, you can customize your Amazon EMR AMI to use the latest Amazon Linux AMI.
- Console support to create a security configuration that specifies the AWS Ranger integration option is currently not supported in the GovCloud Region. Security configuration can be done using the CLI. See Create the EMR Security Configuration in the *Amazon EMR Management Guide*.
- When AtRestEncryption or HDFS encryption is enabled on a cluster that uses Amazon EMR 5.31.0 or 5.32.0, Hive queries result in the following runtime exception.

```
TaskAttempt 3 failed, info=[Error: Error while running task ( failure ) :
 attempt_1604112648850_0001_1_01_000000_3:java.lang.RuntimeException:
 java.lang.RuntimeException: Hive Runtime Error while closing
 operators: java.io.IOException: java.util.ServiceConfigurationError:
 org.apache.hadoop.security.token.TokenIdentifier: Provider
 org.apache.hadoop.hbase.security.token.AuthenticationTokenIdentifier not found
```

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&'()*+,-`. For more information, see UTF-8 encoding table and Unicode characters.

As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

5.32.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.13.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.1.0	Conda env for emr notebook which includes jupyter enterprise gateway

Component	Version	Description
emr-s3-dist-cp	2.17.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.6.0	EMR S3Select Connector
emrfs	2.45.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.11.2	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.11.2	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.10.1-amzn-0	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.10.1-amzn-0	HDFS node-level service for storing blocks.

Component	Version	Description
hadoop-hdfs-library	2.10.1-amzn-0	HDFS command-line client and library
hadoop-hdfs-namenode	2.10.1-amzn-0	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	2.10.1-amzn-0	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.10.1-amzn-0	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.10.1-amzn-0	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.10.1-amzn-0	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.10.1-amzn-0	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.10.1-amzn-0	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.10.1-amzn-0	Service for retrieving current and historical information for YARN applications.

Component	Version	Description
hbase-hmaster	1.4.13	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.13	Service for serving one or more HBase regions.
hbase-client	1.4.13	HBase command-line client.
hbase-rest-server	1.4.13	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.13	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.7-amzn-3	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.7-amzn-3	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.7-amzn-3	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.7-amzn-3	Hive command line client.
hive-hbase	2.3.7-amzn-3	Hive-hbase client.
hive-metastore-server	2.3.7-amzn-3	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.

Component	Version	Description
hive-server2	2.3.7-amzn-3	Service for accepting Hive queries as web requests.
hudi	0.6.0-amzn-0	Incremental processing framework to power data pipline at low latency and high efficiency.
hudi-spark	0.6.0-amzn-0	Bundle library for running Spark with Hudi.
hudi-presto	0.6.0-amzn-0	Bundle library for running Presto with Hudi.
hue-server	4.8.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.1.0	Multi-user server for Jupyter notebooks
livy-server	0.7.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.7.0	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.68	MySQL database server.
nvidia-cuda	10.1.243	Nvidia drivers and Cuda toolkit

Component	Version	Description
oozie-client	5.2.0	Oozie command-line client.
oozie-server	5.2.0	Service for accepting Oozie workflow requests.
opencv	4.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.3-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.240.1-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.240.1-amzn-0	Service for executing pieces of a query.
presto-client	0.240.1-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
pig-client	0.17.0	Pig command-line client.
r	3.4.3	The R Project for Statistical Computing

Component	Version	Description
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.7-amzn-0	Spark command-line clients.
spark-history-server	2.4.7-amzn-0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.7-amzn-0	In-memory execution engine for YARN.
spark-yarn-slave	2.4.7-amzn-0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.3.1	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.2	Web-based notebook that enables interactive data analytics.

Component	Version	Description
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

5.32.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

Reconfiguration actions occur when you specify a configuration for instance groups in a running cluster. Amazon EMR only initiates reconfiguration actions for the classifications that you modify. For more information, see Reconfigure an instance group in a running cluster.

emr-5.32.0 classifications

Classifications	Description	Reconfiguration Actions
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.	Restarts the ResourceM anager service.
container-executor	Change values in Hadoop YARN's container-executor.cfg file.	Not available.
container-log4j	Change values in Hadoop YARN's container-log4j.pr operties file.	Not available.
core-site	Change values in Hadoop's core-site.xml file.	Restarts the Hadoop HDFS services Namenode,

Classifications	Description	Reconfiguration Actions
		SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts Hadoop KMS, Ranger KMS, HiveServe r2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Change docker related settings.	Not available.
emrfs-site	Change EMRFS settings.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Resta rts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ er, and TimelineServer. Ad ditionally restarts HBaseRegi onserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore , Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Change flink-conf.yaml settings.	Restarts Flink history server.

Classifications	Description	Reconfiguration Actions
flink-log4j	Change Flink log4j.properties settings.	Restarts Flink history server.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.	Not available.
flink-log4j-cli	Change Flink log4j-cli .properties settings.	Restarts Flink history server.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Change values in Hadoop's log4j.properties file.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
hadoop-ssl-server	Change hadoop ssl server configuration	Not available.
hadoop-ssl-client	Change hadoop ssl client configuration	Not available.
hbase	Amazon EMR-curated settings for Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Change values in HBase's environment.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Change values in HBase's hbase-policy.xml file.	Not available.
hbase-site	Change values in HBase's hbase-site.xml file.	Restarts the HBase services RegionServer, HBaseMast er, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure HDFS encryption zones.	Should not be reconfigured.

Classifications	Description	Reconfiguration Actions
hdfs-site	Change values in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Change values in HCatalog's environment.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.	Not available.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.	Not available.
hive-env	Change values in the Hive environment.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Classifications	Description	Reconfiguration Actions
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2.properties file.	Not available.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.	Not available.
hive-site	Change values in Hive's hive-site.xml file	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file	Not available.
hue-ini	Change values in Hue's ini file	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Change values in the HTTPFS environment.	Restarts Hadoop Httpfs service.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.	Not available.
hadoop-kms-env	Change values in the Hadoop KMS environment.	Restarts Hadoop-KMS service.

Classifications	Description	Reconfiguration Actions
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.	Not available.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Change values in the Hudi environment.	Not available.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_n otebook_config.py file.	Not available.
jupyter-hub-conf	Change values in JupyterHu bs's jupyterhub_config.py file.	Not available.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.	Not available.
jupyter-sparkmagic-conf	Change values in Sparkmagi c's config.json file.	Not available.
livy-conf	Change values in Livy's livy.conf file.	Restarts Livy Server.
livy-env	Change values in the Livy environment.	Restarts Livy Server.
livy-log4j	Change Livy log4j.properties settings.	Restarts Livy Server.
mapred-env	Change values in the MapReduce application's environment.	Restarts Hadoop MapReduce-HistoryServer.

Classifications	Description	Reconfiguration Actions
mapred-site	Change values in the MapReduce application's mapred-site.xml file.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Change values in Oozie's environment.	Restarts Oozie.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.	Restarts Oozie.
oozie-site	Change values in Oozie's oozie-site.xml file.	Restarts Oozie.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hb ase.properties file.	Not available.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.	Not available.
phoenix-log4j	Change values in Phoenix's log4j.properties file.	Restarts Phoenix-QueryServe r.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.pro perties file.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Change values in Pig's pig.properties file.	Restarts Oozie.
pig-log4j	Change values in Pig's log4j.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-log	Change values in Presto's log.properties file.	Restarts Presto-Server.
presto-config	Change values in Presto's config.properties file.	Restarts Presto-Server.
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.	Not available.
presto-env	Change values in Presto's presto-env.sh file.	Restarts Presto-Server.
presto-node	Change values in Presto's node.properties file.	Not available.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.	Not available.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.	Not available.
presto-connector-hive	Change values in Presto's hive.properties file.	Restarts Presto-Server.
presto-connector-jmx	Change values in Presto's jmx.properties file.	Not available.
presto-connector-kafka	Change values in Presto's kafka.properties file.	Not available.
presto-connector-localfile	Change values in Presto's localfile.properties file.	Not available.
presto-connector-memory	Change values in Presto's memory.properties file.	Not available.

Classifications	Description	Reconfiguration Actions
presto-connector-mongodb	Change values in Presto's mongodb.properties file.	Not available.
presto-connector-mysql	Change values in Presto's mysql.properties file.	Not available.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.	Not available.
presto-connector-raptor	Change values in Presto's raptor.properties file.	Not available.
presto-connector-redis	Change values in Presto's redis.properties file.	Not available.
presto-connector-redshift	Change values in Presto's redshift.properties file.	Not available.
presto-connector-tpch	Change values in Presto's tpch.properties file.	Not available.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.	Not available.
ranger-kms-dbks-site	Change values in dbks-site .xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Change values in the Ranger KMS environment.	Restarts Ranger KMS Server.
ranger-kms-log4j	Change values in kms-log4j .properties file of Ranger KMS.	Not available.

Classifications	Description	Reconfiguration Actions
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.	Not available.
recordserver-env	Change values in the EMR RecordServer environment.	Restarts EMR record server.
recordserver-conf	Change values in EMR RecordServer's erver.pro perties file.	Restarts EMR record server.
recordserver-log4j	Change values in EMR RecordServer's log4j.pro perties file.	Restarts EMR record server.
spark	Amazon EMR-curated settings for Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Change values in Spark's spark-defaults.conf file.	Restarts Spark history server and Spark thrift server.
spark-env	Change values in the Spark environment.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Change values in Spark's hive-site.xml file	Not available.
spark-log4j	Change values in Spark's log4j.properties file.	Restarts Spark history server and Spark thrift server.
spark-metrics	Change values in Spark's metrics.properties file.	Restarts Spark history server and Spark thrift server.
sqoop-env	Change values in Sqoop's environment.	Not available.

Classifications	Description	Reconfiguration Actions
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.	Not available.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.	Not available.
tez-site	Change values in Tez's tez-site.xml file.	Restarts Oozie.
yarn-env	Change values in the YARN environment.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Change values in YARN's yarn-site.xml file.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Change values in the Zeppelin environment.	Restarts Zeppelin.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.	Restarts Zookeeper server.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.	Restarts Zookeeper server.

Amazon EMR release 5.31.1

5.31.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.31.1	emr-5.31.0	emr-5.30.2	emr-5.30.1
AWS SDK for Java	1.11.852	1.11.852	1.11.759	1.11.759
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.0	1.11.0	1.10.0	1.10.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13

	emr-5.31.1	emr-5.31.0	emr-5.30.2	emr-5.30.1
HCatalog	2.3.7	2.3.7	2.3.6	2.3.6
Hadoop	2.10.0	2.10.0	2.8.5	2.8.5
Hive	2.3.7	2.3.7	2.3.6	2.3.6
Hudi	0.6.0-amzn-0	0.6.0-amzn-0	0.5.2-incubating	0.5.2-incubating
Hue	4.7.1	4.7.1	4.6.0	4.6.0
Iceberg	-	-	-	-
JupyterEn terpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.238.3	0.238.3	0.232	0.232
Spark	2.4.6	2.4.6	2.4.5	2.4.5
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.31.1	emr-5.31.0	emr-5.30.2	emr-5.30.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

5.31.1 release notes

This is a release to fix issues with Amazon EMR Scaling when it fails to scale up/scale down a cluster successfully or causes application failures.

Changes, Enhancements, and Resolved Issues

- Fixed an issue where scaling requests failed for a large, highly utilized cluster when Amazon EMR on-cluster daemons were running health checking activities, such as gathering YARN node state and HDFS node state. This was happening because on-cluster daemons were not able to communicate the health status data of a node to internal Amazon EMR components.
- Improved EMR on-cluster daemons to correctly track the node states when IP addresses are reused to improve reliability during scaling operations.
- SPARK-29683. Fixed an issue where job failures occurred during cluster scale-down as Spark was assuming all available nodes were deny-listed.
- YARN-9011. Fixed an issue where job failures occurred due to a race condition in YARN decommissioning when cluster tried to scale up or down.
- Fixed issue with step or job failures during cluster scaling by ensuring that the node states are always consistent between the Amazon EMR on-cluster daemons and YARN/HDFS.
- Fixed an issue where cluster operations such as scale down and step submission failed for Amazon EMR clusters enabled with Kerberos authentication. This was because the Amazon EMR on-cluster daemon did not renew the Kerberos ticket, which is required to securely communicate with HDFS/YARN running on the primary node.
- Newer Amazon EMR releases fix the issue with a lower "Max open files" limit on older AL2 in Amazon EMR. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later now include a permanent fix with a higher "Max open files" setting.

- HTTPS is now enabled by default for Amazon Linux repositories. If you are using an Amazon S3 VPCE policy to restrict access to specific buckets, you must add the new Amazon Linux bucket ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` to your policy (replace `$region` with the region where the endpoint is). For more information, see this topic in the AWS discussion forums. Announcement: Amazon Linux 2 now supports the ability to use HTTPS while connecting to package repositories .

Known issues

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&‘()*+,-`. For more information, see UTF-8 encoding table and Unicode characters.

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

5.31.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2

has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.0	Amazon SageMaker Spark SDK
emr-ddb	4.15.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.13.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.15.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.6.0	EMR S3Select Connector
emrfs	2.43.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.11.0	Apache Flink command line client scripts and applications.
flink-jobmanager-config	1.11.0	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.

Component	Version	Description
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.10.0-amzn-0.1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.10.0-amzn-0.1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.10.0-amzn-0.1	HDFS command-line client and library
hadoop-hdfs-namenode	2.10.0-amzn-0.1	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	2.10.0-amzn-0.1	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.10.0-amzn-0.1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.10.0-amzn-0.1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.10.0-amzn-0.1	MapReduce execution engine libraries for running a MapReduce application.

Component	Version	Description
hadoop-yarn-nodemanager	2.10.0-amzn-0.1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.10.0-amzn-0.1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.10.0-amzn-0.1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.13	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.13	Service for serving one or more HBase regions.
hbase-client	1.4.13	HBase command-line client.
hbase-rest-server	1.4.13	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.13	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.7-amzn-1	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.7-amzn-1	Service providing HCatalog, a table and storage management layer for distributed applications.

Component	Version	Description
hcatalog-webhcat-server	2.3.7-amzn-1	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.7-amzn-1	Hive command line client.
hive-hbase	2.3.7-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.7-amzn-1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.7-amzn-1	Service for accepting Hive queries as web requests.
hudi	0.6.0-amzn-0	Incremental processing framework to power data pipline at low latency and high efficiency.
hudi-spark	0.6.0-amzn-0	Bundle library for running Spark with Hudi.
hudi-presto	0.6.0-amzn-0	Bundle library for running Presto with Hudi.
hue-server	4.7.1	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.1.0	Multi-user server for Jupyter notebooks
livy-server	0.7.0-incubating	REST interface for interacting with Apache Spark

Component	Version	Description
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.6.0	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.64+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.2.0	Oozie command-line client.
oozie-server	5.2.0	Service for accepting Oozie workflow requests.
opencv	4.3.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.3-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.238.3-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.

Component	Version	Description
presto-worker	0.238.3-amzn-0	Service for executing pieces of a query.
presto-client	0.238.3-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
pig-client	0.17.0	Pig command-line client.
r	3.4.3	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.6-amzn-0.1	Spark command-line clients.
spark-history-server	2.4.6-amzn-0.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.6-amzn-0.1	In-memory execution engine for YARN.
spark-yarn-slave	2.4.6-amzn-0.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.1.0	TensorFlow open source software library for high performance numerical computation.

Component	Version	Description
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.2	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

5.31.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.31.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.

Classifications	Description
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.

Classifications	Description
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file

Classifications	Description
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
hudi-env	Change values in the Hudi environment.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.

Classifications	Description
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.

Classifications	Description
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.
ranger-kms-dbks-site	Change values in dbks-site.xml file of Ranger KMS.

Classifications	Description
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.
ranger-kms-env	Change values in the Ranger KMS environme nt.
ranger-kms-log4j	Change values in kms-log4j.properties file of Ranger KMS.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.
recordserver-env	Change values in the EMR RecordServer environment.
recordserver-conf	Change values in EMR RecordServer's erver.pro perties file.
recordserver-log4j	Change values in EMR RecordServer's log4j.pro perties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.

Classifications	Description
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.31.0

5.31.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.31.0	emr-5.30.2	emr-5.30.1	emr-5.30.0
AWS SDK for Java	1.11.852	1.11.759	1.11.759	1.11.759
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.0	1.10.0	1.10.0	1.10.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.7	2.3.6	2.3.6	2.3.6
Hadoop	2.10.0	2.8.5	2.8.5	2.8.5
Hive	2.3.7	2.3.6	2.3.6	2.3.6
Hudi	0.6.0-amzn-0	0.5.2-incubating	0.5.2-incubating	0.5.2-incubating
Hue	4.7.1	4.6.0	4.6.0	4.6.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.6.0	1.5.1	1.5.1	1.5.1

	emr-5.31.0	emr-5.30.2	emr-5.30.1	emr-5.30.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.238.3	0.232	0.232	0.232
Spark	2.4.6	2.4.5	2.4.5	2.4.5
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.1.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

5.31.0 release notes

The following release notes include information for Amazon EMR release 5.31.0. Changes are relative to 5.30.1.

Initial release date: Oct 9, 2020

Last updated date: Oct 15, 2020

Upgrades

- Upgraded Amazon Glue connector to version 1.13.0
- Upgraded Amazon SageMaker Spark SDK to version 1.4.0
- Upgraded Amazon Kinesis connector to version 3.5.9

- Upgraded AWS SDK for Java to version 1.11.852
- Upgraded Bigtop-tomcat to version 8.5.56
- Upgraded EMR FS to version 2.43.0
- Upgraded EMR MetricsAndEventsApiGateway Client to version 1.4.0
- Upgraded EMR S3 Dist CP to version 2.15.0
- Upgraded EMR S3 Select to version 1.6.0
- Upgraded Flink to version 1.11.0
- Upgraded Hadoop to version 2.10.0
- Upgraded Hive to version 2.3.7
- Upgraded Hudi to version 0.6.0
- Upgraded Hue to version 4.7.1
- Upgraded JupyterHub to version 1.1.0
- Upgraded Mxnet to version 1.6.0
- Upgraded OpenCV to version 4.3.0
- Upgraded Presto to version 0.238.3
- Upgraded TensorFlow to version 2.1.0

Changes, enhancements, and resolved issues

- This is a release to fix issues with Amazon EMR Scaling when it fails to scale up/scale down a cluster successfully or causes application failures.
- Fixed an issue where scaling requests failed for a large, highly utilized cluster when Amazon EMR on-cluster daemons were running health checking activities, such as gathering YARN node state and HDFS node state. This was happening because on-cluster daemons were not able to communicate the health status data of a node to internal Amazon EMR components.
- Improved EMR on-cluster daemons to correctly track the node states when IP addresses are reused to improve reliability during scaling operations.
- SPARK-29683. Fixed an issue where job failures occurred during cluster scale-down as Spark was assuming all available nodes were deny-listed.
- YARN-9011. Fixed an issue where job failures occurred due to a race condition in YARN decommissioning when cluster tried to scale up or down.

- Fixed issue with step or job failures during cluster scaling by ensuring that the node states are always consistent between the Amazon EMR on-cluster daemons and YARN/HDFS.
- Fixed an issue where cluster operations such as scale down and step submission failed for Amazon EMR clusters enabled with Kerberos authentication. This was because the Amazon EMR on-cluster daemon did not renew the Kerberos ticket, which is required to securely communicate with HDFS/YARN running on the primary node.
- Newer Amazon EMR releases fix the issue with a lower "Max open files" limit on older AL2 in Amazon EMR. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later now include a permanent fix with a higher "Max open files" setting.
- Hive column statistics are supported for Amazon EMR versions 5.31.0 and later.
- Upgraded component versions.
- EMRFS S3EC V2 Support in Amazon EMR 5.31.0. In S3 Java SDK releases 1.11.837 and later, encryption client Version 2 (S3EC V2) has been introduced with various security enhancements. For more information, see the following:
  - S3 blog post: Updates to the Amazon S3 encryption client.
  - AWS SDK for Java Developer Guide: Migrate encryption and decryption clients to V2.
  - EMR Management Guide: Amazon S3 client-side encryption.

  Encryption Client V1 is still available in the SDK for backward compatibility.

New features

- **Lower "Max open files" limit on older AL2 [fixed in newer releases].** Amazon EMR releases: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0, and emr-6.2.0 are based on older versions ofAmazon Linux 2 (AL2), which have a lower ulimit setting for "Max open files" when Amazon EMR clusters are created with the default AMI. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later include a permanent fix with a higher "Max open files" setting. Releases with the lower open file limit causes a "Too many open files" error when submitting Spark job. In the impacted releases, the Amazon EMR default AMI has a default ulimit setting of 4096 for "Max open files," which is lower than the 65536 file limit in the latestAmazon Linux 2 AMI. The lower ulimit setting for "Max open files" causes Spark job failure when the Spark driver and executor try to open more than 4096 files. To fix the issue, Amazon EMR has a bootstrap action (BA) script that adjusts the ulimit setting at cluster creation.

If you are using an older Amazon EMR version that doesn't have the permanent fix for this issue, the following workaround lets you to explicitly set the instance-controller ulimit to a maximum of 65536 files.

Explicitly set a ulimit from the command line

1. Edit `/etc/systemd/system/instance-controller.service` to add the following parameters to Service section.

   `LimitNOFILE=65536`

   `LimitNPROC=65536`

2. Restart InstanceController

   `$ sudo systemctl daemon-reload`

   `$ sudo systemctl restart instance-controller`

Set a ulimit using bootstrap action (BA)

You can also use a bootstrap action (BA) script to configure the instance-controller ulimit to 65536 files at cluster creation.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
```

```
sudo systemctl daemon-reload
```

- With Amazon EMR 5.31.0, you can launch a cluster that integrates with Lake Formation. This integration provides fine-grained, column-level data filtering to databases and tables in the AWS Glue Data Catalog. It also enables federated single sign-on to EMR Notebooks or Apache Zeppelin from an enterprise identity system. For more information, see Integrating Amazon EMR with AWS Lake Formation in the *Amazon EMR Management Guide*.

  Amazon EMR with Lake Formation is currently available in 16 AWS Regions: US East (Ohio and N. Virginia), US West (N. California and Oregon), Asia Pacific (Mumbai, Seoul, Singapore, Sydney, and Tokyo), Canada (Central), Europe (Frankfurt, Ireland, London, Paris, and Stockholm), South America (São Paulo).

Known issues

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as hadoop user to the lead primary node of the EMR cluster with multiple primary nodes.
  - Run the following command to renew Kerberos ticket for hadoop user.

    ```
    kinit -kt <keytab_file> <principal>
    ```

    Typically, the keytab file is located at /etc/hadoop.keytab and the principal is in the form of hadoop/<hostname>@<REALM>.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

- When AtRestEncryption or HDFS encryption is enabled on a cluster that uses Amazon EMR 5.31.0 or 5.32.0, Hive queries result in the following runtime exception.

```
TaskAttempt 3 failed, info=[Error: Error while running task ( failure ) :
 attempt_1604112648850_0001_1_01_000000_3:java.lang.RuntimeException:
 java.lang.RuntimeException: Hive Runtime Error while closing
 operators: java.io.IOException: java.util.ServiceConfigurationError:
 org.apache.hadoop.security.token.TokenIdentifier: Provider
 org.apache.hadoop.hbase.security.token.AuthenticationTokenIdentifier not found
```

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&'()*+,-`. For more information, see [UTF-8 encoding table and Unicode characters](#).

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

5.31.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system

processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.4.0	Amazon SageMaker Spark SDK
emr-ddb	4.15.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.13.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.15.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.6.0	EMR S3Select Connector
emrfs	2.43.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.11.0	Apache Flink command line client scripts and applications.

Component	Version	Description
flink-jobmanager-config	1.11.0	Managing resources on EMR nodes for Apache Flink JobManager.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.10.0-amzn-0	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.10.0-amzn-0	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.10.0-amzn-0	HDFS command-line client and library
hadoop-hdfs-namenode	2.10.0-amzn-0	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	2.10.0-amzn-0	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.10.0-amzn-0	HTTP endpoint for HDFS operations.

Component	Version	Description
hadoop-kms-server	2.10.0-amzn-0	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.10.0-amzn-0	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.10.0-amzn-0	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.10.0-amzn-0	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.10.0-amzn-0	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.13	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.13	Service for serving one or more HBase regions.
hbase-client	1.4.13	HBase command-line client.
hbase-rest-server	1.4.13	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.13	Service providing a Thrift endpoint to HBase.

Component	Version	Description
hcatalog-client	2.3.7-amzn-1	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.7-amzn-1	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.7-amzn-1	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.7-amzn-1	Hive command line client.
hive-hbase	2.3.7-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.7-amzn-1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.7-amzn-1	Service for accepting Hive queries as web requests.
hudi	0.6.0-amzn-0	Incremental processing framework to power data pipline at low latency and high efficiency.
hudi-spark	0.6.0-amzn-0	Bundle library for running Spark with Hudi.
hudi-presto	0.6.0-amzn-0	Bundle library for running Presto with Hudi.

Component	Version	Description
hue-server	4.7.1	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.1.0	Multi-user server for Jupyter notebooks
livy-server	0.7.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.6.0	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.64	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.2.0	Oozie command-line client.
oozie-server	5.2.0	Service for accepting Oozie workflow requests.
opencv	4.3.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	The phoenix libraries for server and client

Component	Version	Description
phoenix-query-server	4.14.3-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.238.3-amzn-0	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.238.3-amzn-0	Service for executing pieces of a query.
presto-client	0.238.3-amzn-0	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
pig-client	0.17.0	Pig command-line client.
r	3.4.3	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.6-amzn-0	Spark command-line clients.
spark-history-server	2.4.6-amzn-0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.6-amzn-0	In-memory execution engine for YARN.

Component	Version	Description
spark-yarn-slave	2.4.6-amzn-0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	2.1.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.2	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

5.31.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.31.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.

Classifications	Description
hbase-metrics	Change values in HBase's hadoop-metrics2-hbase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.

Classifications	Description
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2- site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
hudi-env	Change values in the Hudi environment.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.

Classifications	Description
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.

Classifications	Description
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.

Classifications	Description
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.
ranger-kms-dbks-site	Change values in dbks-site.xml file of Ranger KMS.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.
ranger-kms-env	Change values in the Ranger KMS environme nt.
ranger-kms-log4j	Change values in kms-log4j.properties file of Ranger KMS.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.
recordserver-env	Change values in the EMR RecordServer environment.
recordserver-conf	Change values in EMR RecordServer's erver.pro perties file.
recordserver-log4j	Change values in EMR RecordServer's log4j.pro perties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.

Classifications	Description
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.30.2

5.30.2 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.30.2	emr-5.30.1	emr-5.30.0	emr-5.29.0
AWS SDK for Java	1.11.759	1.11.759	1.11.759	1.11.682
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.10.0	1.10.0	1.10.0	1.9.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.10
HCatalog	2.3.6	2.3.6	2.3.6	2.3.6
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.6	2.3.6
Hudi	0.5.2-incubating	0.5.2-incubating	0.5.2-incubating	0.5.0-incubating

	emr-5.30.2	emr-5.30.1	emr-5.30.0	emr-5.29.0
Hue	4.6.0	4.6.0	4.6.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.0.0
Livy	0.7.0	0.7.0	0.7.0	0.6.0
MXNet	1.5.1	1.5.1	1.5.1	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.1.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.232	0.232	0.232	0.227
Spark	2.4.5	2.4.5	2.4.5	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

5.30.2 release notes

This is a release to fix issues with Amazon EMR Scaling when it fails to scale up/scale down a cluster successfully or causes application failures.

Changes, Enhancements, and Resolved Issues

- Fixed an issue where scaling requests failed for a large, highly utilized cluster when Amazon EMR on-cluster daemons were running health checking activities, such as gathering YARN node state and HDFS node state. This was happening because on-cluster daemons were not able to communicate the health status data of a node to internal Amazon EMR components.
- Improved EMR on-cluster daemons to correctly track the node states when IP addresses are reused to improve reliability during scaling operations.
- SPARK-29683. Fixed an issue where job failures occurred during cluster scale-down as Spark was assuming all available nodes were deny-listed.
- YARN-9011. Fixed an issue where job failures occurred due to a race condition in YARN decommissioning when cluster tried to scale up or down.
- Fixed issue with step or job failures during cluster scaling by ensuring that the node states are always consistent between the Amazon EMR on-cluster daemons and YARN/HDFS.
- Fixed an issue where cluster operations such as scale down and step submission failed for Amazon EMR clusters enabled with Kerberos authentication. This was because the Amazon EMR on-cluster daemon did not renew the Kerberos ticket, which is required to securely communicate with HDFS/YARN running on the primary node.
- Newer Amazon EMR releases fix the issue with a lower "Max open files" limit on older AL2 in Amazon EMR. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later now include a permanent fix with a higher "Max open files" setting.
- HTTPS is now enabled by default for Amazon Linux repositories. If you are using an Amazon S3 VPCE policy to restrict access to specific buckets, you must add the new Amazon Linux bucket ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` to your policy (replace `$region` with the region where the endpoint is). For more information, see this topic in the AWS discussion forums. Announcement: Amazon Linux 2 now supports the ability to use HTTPS while connecting to package repositories .

Known issues

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&‘()*+,-`. For more information, see UTF-8 encoding table and Unicode characters.

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

5.30.2 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.3.0	Amazon SageMaker Spark SDK

Component	Version	Description
emr-ddb	4.14.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.13.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.14.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.5.0	EMR S3Select Connector
emrfs	2.40.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.10.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.

Component	Version	Description
hadoop-client	2.8.5-amzn-6.1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-6.1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.5-amzn-6.1	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.5-amzn-6.1	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	2.8.5-amzn-6.1	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.8.5-amzn-6.1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.5-amzn-6.1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.5-amzn-6.1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-6.1	YARN service for managing containers on an individual node.
hadoop-yarn-resourcemanager	2.8.5-amzn-6.1	YARN service for allocating and managing cluster resources and distributed applications.

Component	Version	Description
hadoop-yarn-timeline-server	2.8.5-amzn-6.1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.13	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.13	Service for serving one or more HBase regions.
hbase-client	1.4.13	HBase command-line client.
hbase-rest-server	1.4.13	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.13	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.6-amzn-2	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.6-amzn-2	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.6-amzn-2	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.6-amzn-2	Hive command line client.
hive-hbase	2.3.6-amzn-2	Hive-hbase client.

Component	Version	Description
hive-metastore-server	2.3.6-amzn-2	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.6-amzn-2	Service for accepting Hive queries as web requests.
hudi	0.5.2-incubating	Incremental processing framework to power data pipline at low latency and high efficiency.
hudi-presto	0.5.2-incubating	Bundle library for running Presto with Hudi.
hue-server	4.6.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.1.0	Multi-user server for Jupyter notebooks
livy-server	0.7.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.5.1	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.64+	MySQL database server.

Component	Version	Description
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.2.0	Oozie command-line client.
oozie-server	5.2.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.3-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.232	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.232	Service for executing pieces of a query.
presto-client	0.232	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
pig-client	0.17.0	Pig command-line client.

Component	Version	Description
r	3.4.3	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.5-amzn-0.1	Spark command-line clients.
spark-history-server	2.4.5-amzn-0.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.5-amzn-0.1	In-memory execution engine for YARN.
spark-yarn-slave	2.4.5-amzn-0.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	1.14.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.2	Web-based notebook that enables interactive data analytics.

Component	Version	Description
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

5.30.2 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.30.2 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.

Classifications	Description
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.

Classifications	Description
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.

Classifications	Description
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
hudi-env	Change values in the Hudi environment.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.

Classifications	Description
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.

Classifications	Description
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.
ranger-kms-dbks-site	Change values in dbks-site.xml file of Ranger KMS.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.
ranger-kms-env	Change values in the Ranger KMS environme nt.

Classifications	Description
ranger-kms-log4j	Change values in kms-log4j.properties file of Ranger KMS.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.
recordserver-env	Change values in the EMR RecordServer environment.
recordserver-conf	Change values in EMR RecordServer's erver.pro perties file.
recordserver-log4j	Change values in EMR RecordServer's log4j.pro perties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.

Classifications	Description
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.30.1

5.30.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.30.1	emr-5.30.0	emr-5.29.0	emr-5.28.1
AWS SDK for Java	1.11.759	1.11.759	1.11.682	1.11.659

	emr-5.30.1	emr-5.30.0	emr-5.29.0	emr-5.28.1
Python	2.7, 3.7	2.7, 3.7	2.7, 3.6	2.7, 3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.10.0	1.10.0	1.9.1	1.9.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.10	1.4.10
HCatalog	2.3.6	2.3.6	2.3.6	2.3.6
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.6	2.3.6
Hudi	0.5.2-incubating	0.5.2-incubating	0.5.0-incubating	0.5.0-incubating
Hue	4.6.0	4.6.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0
MXNet	1.5.1	1.5.1	1.5.1	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.1.0	5.1.0

	emr-5.30.1	emr-5.30.0	emr-5.29.0	emr-5.28.1
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.232	0.232	0.227	0.227
Spark	2.4.5	2.4.5	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

5.30.1 release notes

The following release notes include information for Amazon EMR release 5.30.1. Changes are relative to 5.30.0.

Initial release date: June 30, 2020

Last updated date: August 24, 2020

Changes, enhancements, and resolved issues

- Newer Amazon EMR releases fix the issue with a lower "Max open files" limit on older AL2 in Amazon EMR. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later now include a permanent fix with a higher "Max open files" setting.
- Fixed issue where instance controller process spawned infinite number of processes.
- Fixed issue where Hue was unable to run an Hive query, showing a "database is locked" message and preventing the execution of queries.

- Fixed a Spark issue to enable more tasks to run concurrently on the EMR cluster.
- Fixed a Jupyter notebook issue causing a "too many files open error" in the Jupyter server.
- Fixed an issue with cluster start times.

New features

- Tez UI and YARN timeline server persistent application interfaces are available with Amazon EMR versions 6.x, and EMR version 5.30.1 and later. One-click link access to persistent application history lets you quickly access job history without setting up a web proxy through an SSH connection. Logs for active and terminated clusters are available for 30 days after the application ends. For more information, see View Persistent Application User Interfaces in the *Amazon EMR Management Guide*.
- EMR Notebook execution APIs are available to execute EMR notebooks via a script or command line. The ability to start, stop, list, and describe EMR notebook executions without the AWS console enables you programmatically control an EMR notebook. Using a parameterized notebook cell, you can pass different parameter values to a notebook without having to create a copy of the notebook for each new set of paramter values. See EMR API Actions. For sample code, see Sample commands to execute EMR Notebooks programmatically.

Known issues

- **Lower "Max open files" limit on older AL2 [fixed in newer releases].** Amazon EMR releases: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0, and emr-6.2.0 are based on older versions ofAmazon Linux 2 (AL2), which have a lower ulimit setting for "Max open files" when Amazon EMR clusters are created with the default AMI. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later include a permanent fix with a higher "Max open files" setting. Releases with the lower open file limit causes a "Too many open files" error when submitting Spark job. In the impacted releases, the Amazon EMR default AMI has a default ulimit setting of 4096 for "Max open files," which is lower than the 65536 file limit in the latestAmazon Linux 2 AMI. The lower ulimit setting for "Max open files" causes Spark job failure when the Spark driver and executor try to open more than 4096 files. To fix the issue, Amazon EMR has a bootstrap action (BA) script that adjusts the ulimit setting at cluster creation.

  If you are using an older Amazon EMR version that doesn't have the permanent fix for this issue, the following workaround lets you to explicitly set the instance-controller ulimit to a maximum of 65536 files.

Explicitly set a ulimit from the command line

1. Edit `/etc/systemd/system/instance-controller.service` to add the following parameters to Service section.

   `LimitNOFILE=65536`

   `LimitNPROC=65536`

2. Restart InstanceController

   `$ sudo systemctl daemon-reload`

   `$ sudo systemctl restart instance-controller`

Set a ulimit using bootstrap action (BA)

You can also use a bootstrap action (BA) script to configure the instance-controller ulimit to 65536 files at cluster creation.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- **EMR Notebooks**

The feature that allows you to install kernels and additional Python libraries on the cluster primary node is disabled by default on EMR version 5.30.1. For more information about this feature, see Installing Kernels and Python Libraries on a Cluster Primary Node.

To enable the feature, do the following:

1. Make sure that the permissions policy attached to the service role for EMR Notebooks allows the following action:

   ```
   elasticmapreduce:ListSteps
   ```

   For more information, see Service Role for EMR Notebooks.
2. Use the AWS CLI to run a step on the cluster that sets up EMR Notebooks as shown in the following example. Replace *us-east-1* with the Region in which your cluster resides. For more information, see Adding Steps to a Cluster Using the AWS CLI.

   ```
   aws emr add-steps  --cluster-id MyClusterID --steps
    Type=CUSTOM_JAR,Name=EMRNotebooksSetup,ActionOnFailure=CONTINUE,Jar=s3://us-
   east-1.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://
   awssupportdatasvcs.com/bootstrap-actions/EMRNotebooksSetup/emr-notebooks-setup.sh"]
   ```

- **Managed scaling**

  Managed scaling operations on 5.30.0 and 5.30.1 clusters without Presto installed may cause application failures or cause a uniform instance group or instance fleet to stay in the ARRESTED state, particularly when a scale down operation is followed quickly by a scale up operation.

  As a workaround, choose Presto as an application to install when you create a cluster with Amazon EMR releases 5.30.0 and 5.30.1, even if your job does not require Presto.
- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

- SSH as hadoop user to the lead primary node of the EMR cluster with multiple primary nodes.
- Run the following command to renew Kerberos ticket for hadoop user.

```
kinit -kt <keytab_file> <principal>
```

  Typically, the keytab file is located at /etc/hadoop.keytab and the principal is in the form of hadoop/<hostname>@<REALM>.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&‘()*+,-`. For more information, see UTF-8 encoding table and Unicode characters.

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

5.30.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with emr or aws. Big-data application packages in the

most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
aws-sagemaker-spark-sdk	1.3.0	Amazon SageMaker Spark SDK
emr-ddb	4.14.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.13.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.14.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.5.0	EMR S3Select Connector
emrfs	2.40.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.10.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati

Component	Version	Description
		ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.5-amzn-6	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-6	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.5-amzn-6	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.5-amzn-6	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	2.8.5-amzn-6	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.8.5-amzn-6	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.5-amzn-6	Cryptographic key management server based on Hadoop's KeyProvider API.

Component	Version	Description
hadoop-mapred	2.8.5-amzn-6	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-6	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.5-amzn-6	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.5-amzn-6	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.13	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.13	Service for serving one or more HBase regions.
hbase-client	1.4.13	HBase command-line client.
hbase-rest-server	1.4.13	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.13	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.6-amzn-2	The 'hcat' command line client for manipulating hcatalog-server.

Component	Version	Description
hcatalog-server	2.3.6-amzn-2	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.6-amzn-2	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.6-amzn-2	Hive command line client.
hive-hbase	2.3.6-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.6-amzn-2	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.6-amzn-2	Service for accepting Hive queries as web requests.
hudi	0.5.2-incubating	Incremental processing framework to power data pipline at low latency and high efficiency.
hudi-presto	0.5.2-incubating	Bundle library for running Presto with Hudi.
hue-server	4.6.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.1.0	Multi-user server for Jupyter notebooks

Component	Version	Description
livy-server	0.7.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.5.1	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.64	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.2.0	Oozie command-line client.
oozie-server	5.2.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.3-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API

Component	Version	Description
presto-coordinator	0.232	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.232	Service for executing pieces of a query.
presto-client	0.232	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
pig-client	0.17.0	Pig command-line client.
r	3.4.3	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.5-amzn-0	Spark command-line clients.
spark-history-server	2.4.5-amzn-0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.5-amzn-0	In-memory execution engine for YARN.
spark-yarn-slave	2.4.5-amzn-0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.

Component	Version	Description
tensorflow	1.14.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.2	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

5.30.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.30.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.

Classifications	Description
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.

Classifications	Description
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.

Classifications	Description
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
hudi-env	Change values in the Hudi environment.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.

Classifications	Description
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.

Classifications	Description
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.

Classifications	Description
ranger-kms-dbks-site	Change values in dbks-site.xml file of Ranger KMS.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.
ranger-kms-env	Change values in the Ranger KMS environme nt.
ranger-kms-log4j	Change values in kms-log4j.properties file of Ranger KMS.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.
recordserver-env	Change values in the EMR RecordServer environment.
recordserver-conf	Change values in EMR RecordServer's erver.pro perties file.
recordserver-log4j	Change values in EMR RecordServer's log4j.pro perties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.

Classifications	Description
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.30.0

5.30.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases

- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.30.0	emr-5.29.0	emr-5.28.1	emr-5.28.0
AWS SDK for Java	1.11.759	1.11.682	1.11.659	1.11.659
Python	2.7, 3.7	2.7, 3.6	2.7, 3.6	2.7, 3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.10.0	1.9.1	1.9.0	1.9.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.10	1.4.10	1.4.10
HCatalog	2.3.6	2.3.6	2.3.6	2.3.6
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.6	2.3.6
Hudi	0.5.2-incubating	0.5.0-incubating	0.5.0-incubating	0.5.0-incubating
Hue	4.6.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.0.0	1.0.0	1.0.0

	emr-5.30.0	emr-5.29.0	emr-5.28.1	emr-5.28.0
Livy	0.7.0	0.6.0	0.6.0	0.6.0
MXNet	1.5.1	1.5.1	1.5.1	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.232	0.227	0.227	0.227
Spark	2.4.5	2.4.4	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

5.30.0 release notes

The following release notes include information for Amazon EMR release 5.30.0. Changes are relative to 5.29.0.

Initial release date: May 13, 2020

Last updated date: June 25, 2020

Upgrades

- Upgraded AWS SDK for Java to version 1.11.759
- Upgraded Amazon SageMaker Spark SDK to version 1.3.0
- Upgraded EMR Record Server to version 1.6.0
- Upgraded Flink to version 1.10.0
- Upgraded Ganglia to version 3.7.2
- Upgraded HBase to version 1.4.13
- Upgraded Hudi to version 0.5.2-incubating
- Upgraded Hue to version 4.6.0
- Upgraded JupyterHub to version 1.1.0
- Upgraded Livy to version 0.7.0-incubating
- Upgraded Oozie to version 5.2.0
- Upgraded Presto to version 0.232
- Upgraded Spark to version 2.4.5
- Upgraded Connectors and drivers: Amazon Glue Connector 1.12.0; Amazon Kinesis Connector 3.5.0; EMR DynamoDB Connector 4.14.0

New features

- **EMR Notebooks** – When used with EMR clusters created using 5.30.0, EMR notebook kernels run on cluster. This improves notebook performance and allows you to install and customize kernels. You can also install Python libraries on the cluster primary node. For more information, see Installing and Using Kernels and Libraries in the *EMR Management Guide*.
- **Managed Scaling** – With Amazon EMR version 5.30.0 and later, you can enable EMR managed scaling to automatically increase or decrease the number of instances or units in your cluster based on workload. Amazon EMR continuously evaluates cluster metrics to make scaling decisions that optimize your clusters for cost and speed. For more information, see Scaling Cluster Resources in the *Amazon EMR Management Guide*.
- **Encrypt log files stored in Amazon S3** – With Amazon EMR version 5.30.0 and later, you can encrypt log files stored in Amazon S3 with an AWS KMS customer managed key. For more information, see Encrypt log files stored in Amazon S3 in the *Amazon EMR Management Guide*.

- **Amazon Linux 2 support –** In EMR version 5.30.0 and later, EMR usesAmazon Linux 2 OS. New custom AMIs (Amazon Machine Image) must be based on theAmazon Linux 2 AMI. For more information, see Using a Custom AMI.
- **Presto Graceful Auto Scale –** EMR clusters using 5.30.0 can be set with an auto scaling timeout period that gives Presto tasks time to finish running before their node is decommissioned. For more information, see Using Presto automatic scaling with Graceful Decommission.
- **Fleet Instance creation with new allocation strategy option –** A new allocation strategy option is available in EMR version 5.12.1 and later. It offers faster cluster provisioning, more accurate spot allocation, and less spot instance interruption. Updates to non-default EMR service roles are required. See Configure Instance Fleets.
- **sudo systemctl stop and sudo systemctl start commands –** In EMR version 5.30.0 and later, which useAmazon Linux 2 OS, EMR uses `sudo systemctl stop` and `sudo systemctl start` commands to restart services. For more information, see How do I restart a service in Amazon EMR?.

Changes, enhancements, and resolved issues

- EMR version 5.30.0 doesn't install Ganglia by default. You can explicitly select Ganglia to install when you create a cluster.
- Spark performance optimizations.
- Presto performance optimizations.
- Python 3 is the default for Amazon EMR version 5.30.0 and later.
- The default managed security group for service access in private subnets has been updated with new rules. If you use a custom security group for service access, you must include the same rules as the default managed security group. For more information, see Amazon EMR-Managed Security Group for Service Access (Private Subnets). If you use a custom service role for Amazon EMR, you must grant permission to `ec2:describeSecurityGroups` so that EMR can validate if the security groups are correctly created. If you use the `EMR_DefaultRole`, this permission is already included in the default managed policy.

Known issues

- **Lower "Max open files" limit on older AL2 [fixed in newer releases].** Amazon EMR releases: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0, and emr-6.2.0 are based on older versions ofAmazon Linux 2 (AL2), which have a lower ulimit setting for "Max open files" when

Amazon EMR clusters are created with the default AMI. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later include a permanent fix with a higher "Max open files" setting. Releases with the lower open file limit causes a "Too many open files" error when submitting Spark job. In the impacted releases, the Amazon EMR default AMI has a default ulimit setting of 4096 for "Max open files," which is lower than the 65536 file limit in the latestAmazon Linux 2 AMI. The lower ulimit setting for "Max open files" causes Spark job failure when the Spark driver and executor try to open more than 4096 files. To fix the issue, Amazon EMR has a bootstrap action (BA) script that adjusts the ulimit setting at cluster creation.

If you are using an older Amazon EMR version that doesn't have the permanent fix for this issue, the following workaround lets you to explicitly set the instance-controller ulimit to a maximum of 65536 files.

Explicitly set a ulimit from the command line

1. Edit `/etc/systemd/system/instance-controller.service` to add the following parameters to Service section.

   `LimitNOFILE=65536`

   `LimitNPROC=65536`
2. Restart InstanceController

   `$ sudo systemctl daemon-reload`

   `$ sudo systemctl restart instance-controller`

Set a ulimit using bootstrap action (BA)

You can also use a bootstrap action (BA) script to configure the instance-controller ulimit to 65536 files at cluster creation.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
```

```
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- **Managed scaling**

  Managed scaling operations on 5.30.0 and 5.30.1 clusters without Presto installed may cause application failures or cause a uniform instance group or instance fleet to stay in the `ARRESTED` state, particularly when a scale down operation is followed quickly by a scale up operation.

  As a workaround, choose Presto as an application to install when you create a cluster with Amazon EMR releases 5.30.0 and 5.30.1, even if your job does not require Presto.

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as `hadoop` user to the lead primary node of the EMR cluster with multiple primary nodes.
  - Run the following command to renew Kerberos ticket for `hadoop` user.

    ```
    kinit -kt <keytab_file> <principal>
    ```

    Typically, the keytab file is located at `/etc/hadoop.keytab` and the principal is in the form of `hadoop/<hostname>@<REALM>`.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

- The default database engine for Hue 4.6.0 is SQLite, which causes issues when you try to use Hue with an external database. To fix this, set `engine` in your `hue-ini` configuration classification to `mysql`. This issue has been fixed in Amazon EMR version 5.30.1.
- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&‘()*+,-`. For more information, see [UTF-8 encoding table and Unicode characters](#).

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

5.30.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0.

For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.3.0	Amazon SageMaker Spark SDK
emr-ddb	4.14.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.13.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.5.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-notebook-env	1.0.0	Conda env for emr notebook
emr-s3-dist-cp	2.14.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.5.0	EMR S3Select Connector
emrfs	2.40.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.10.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.

Component	Version	Description
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.5-amzn-6	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-6	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.5-amzn-6	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.5-amzn-6	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	2.8.5-amzn-6	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.8.5-amzn-6	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.5-amzn-6	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.5-amzn-6	MapReduce execution engine libraries for running a MapReduce application.

Component	Version	Description
hadoop-yarn-nodemanager	2.8.5-amzn-6	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.5-amzn-6	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.5-amzn-6	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.13	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.13	Service for serving one or more HBase regions.
hbase-client	1.4.13	HBase command-line client.
hbase-rest-server	1.4.13	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.13	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.6-amzn-2	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.6-amzn-2	Service providing HCatalog, a table and storage management layer for distributed applications.

Component	Version	Description
hcatalog-webhcat-server	2.3.6-amzn-2	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.6-amzn-2	Hive command line client.
hive-hbase	2.3.6-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.6-amzn-2	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.6-amzn-2	Service for accepting Hive queries as web requests.
hudi	0.5.2-incubating	Incremental processing framework to power data pipline at low latency and high efficiency.
hudi-presto	0.5.2-incubating	Bundle library for running Presto with Hudi.
hue-server	4.6.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.1.0	Multi-user server for Jupyter notebooks
livy-server	0.7.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server

Component	Version	Description
mahout-client	0.13.0	Library for machine learning.
mxnet	1.5.1	A flexible, scalable, and efficient library for deep learning.
mariadb-server	5.5.64	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.2.0	Oozie command-line client.
oozie-server	5.2.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.3-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.232	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.232	Service for executing pieces of a query.

Component	Version	Description
presto-client	0.232	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
pig-client	0.17.0	Pig command-line client.
r	3.4.3	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.5-amzn-0	Spark command-line clients.
spark-history-server	2.4.5-amzn-0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.5-amzn-0	In-memory execution engine for YARN.
spark-yarn-slave	2.4.5-amzn-0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	1.14.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.

Component	Version	Description
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.2	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

5.30.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.30.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.

Classifications	Description
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.

Classifications	Description
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.

Classifications	Description
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
hudi-env	Change values in the Hudi environment.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.

Classifications	Description
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.

Classifications	Description
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.
ranger-kms-dbks-site	Change values in dbks-site.xml file of Ranger KMS.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.

Classifications	Description
ranger-kms-env	Change values in the Ranger KMS environme nt.
ranger-kms-log4j	Change values in kms-log4j.properties file of Ranger KMS.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.
recordserver-env	Change values in the EMR RecordServer environment.
recordserver-conf	Change values in EMR RecordServer's erver.pro perties file.
recordserver-log4j	Change values in EMR RecordServer's log4j.pro perties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.

Classifications	Description
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.29.0

5.29.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.29.0	emr-5.28.1	emr-5.28.0	emr-5.27.1
AWS SDK for Java	1.11.682	1.11.659	1.11.659	1.11.615
Python	2.7, 3.6	2.7, 3.6	2.7, 3.6	2.7, 3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.9.1	1.9.0	1.9.0	1.8.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.10	1.4.10	1.4.10
HCatalog	2.3.6	2.3.6	2.3.6	2.3.5
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.6	2.3.5
Hudi	0.5.0-incubating	0.5.0-incubating	0.5.0-incubating	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	1.0.0	1.0.0	1.0.0
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.5.1	1.5.1	1.5.1	1.4.0

	emr-5.29.0	emr-5.28.1	emr-5.28.0	emr-5.27.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.227	0.227	0.227	0.224
Spark	2.4.4	2.4.4	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

5.29.0 release notes

The following release notes include information for Amazon EMR release 5.29.0. Changes are relative to 5.28.1.

Initial release date: Jan 17, 2020

Upgrades

- Upgraded AWS SDK for Java to version 1.11.682
- Upgraded Hive to version 2.3.6
- Upgraded Flink to version 1.9.1
- Upgraded EmrFS to version 2.38.0

- Upgraded EMR DynamoDB Connector to version 4.13.0

Changes, enhancements, and resolved issues

- Spark
  - Spark performance optimizations.
- EMRFS
  - Management Guide updates to emrfs-site.xml default settings for consistent view.

Known issues

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as `hadoop` user to the lead primary node of the EMR cluster with multiple primary nodes.
  - Run the following command to renew Kerberos ticket for `hadoop` user.

    ```
    kinit -kt <keytab_file> <principal>
    ```

    Typically, the keytab file is located at `/etc/hadoop.keytab` and the principal is in the form of `hadoop/<hostname>@<REALM>`.

  > **Note**
  >
  > This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

5.29.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.2.6	Amazon SageMaker Spark SDK
emr-ddb	4.13.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.12.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.13.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.4.0	EMR S3Select Connector
emrfs	2.38.0	Amazon S3 connector for Hadoop ecosystem applicati ons.

Component	Version	Description
flink-client	1.9.1	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.5-amzn-5	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-5	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.5-amzn-5	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.5-amzn-5	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	2.8.5-amzn-5	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.8.5-amzn-5	HTTP endpoint for HDFS operations.

Component	Version	Description
hadoop-kms-server	2.8.5-amzn-5	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.5-amzn-5	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-5	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.5-amzn-5	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.5-amzn-5	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.10	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.10	Service for serving one or more HBase regions.
hbase-client	1.4.10	HBase command-line client.
hbase-rest-server	1.4.10	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.10	Service providing a Thrift endpoint to HBase.

Component	Version	Description
hcatalog-client	2.3.6-amzn-1	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.6-amzn-1	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.6-amzn-1	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.6-amzn-1	Hive command line client.
hive-hbase	2.3.6-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.6-amzn-1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.6-amzn-1	Service for accepting Hive queries as web requests.
hudi	0.5.0-incubating	Incremental processing framework to power data pipline at low latency and high efficiency.
hudi-presto	0.5.0-incubating	Bundle library for running Presto with Hudi.
hue-server	4.4.0	Web application for analyzing data using Hadoop ecosystem applications

Component	Version	Description
jupyterhub	1.0.0	Multi-user server for Jupyter notebooks
livy-server	0.6.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.5.1	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.1.0	Oozie command-line client.
oozie-server	5.1.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.3-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API

Component	Version	Description
presto-coordinator	0.227	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.227	Service for executing pieces of a query.
presto-client	0.227	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.4	Spark command-line clients.
spark-history-server	2.4.4	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.4	In-memory execution engine for YARN.
spark-yarn-slave	2.4.4	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.

Component	Version	Description
tensorflow	1.14.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.2	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

5.29.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.29.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.

Classifications	Description
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.

Classifications	Description
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.

Classifications	Description
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.

Classifications	Description
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.

Classifications	Description
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.

Classifications	Description
ranger-kms-dbks-site	Change values in dbks-site.xml file of Ranger KMS.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.
ranger-kms-env	Change values in the Ranger KMS environme nt.
ranger-kms-log4j	Change values in kms-log4j.properties file of Ranger KMS.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.
recordserver-env	Change values in the EMR RecordServer environment.
recordserver-conf	Change values in EMR RecordServer's erver.pro perties file.
recordserver-log4j	Change values in EMR RecordServer's log4j.pro perties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.

Classifications	Description
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.28.1

5.28.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases

- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.28.1	emr-5.28.0	emr-5.27.1	emr-5.27.0
AWS SDK for Java	1.11.659	1.11.659	1.11.615	1.11.615
Python	2.7, 3.6	2.7, 3.6	2.7, 3.6	2.7, 3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.9.0	1.9.0	1.8.1	1.8.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.10	1.4.10	1.4.10
HCatalog	2.3.6	2.3.6	2.3.5	2.3.5
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.5	2.3.5
Hudi	0.5.0-incubating	0.5.0-incubating	-	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	1.0.0	1.0.0	1.0.0

	emr-5.28.1	emr-5.28.0	emr-5.27.1	emr-5.27.0
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.5.1	1.5.1	1.4.0	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.3	4.14.3	4.14.2	4.14.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.227	0.227	0.224	0.224
Spark	2.4.4	2.4.4	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

5.28.1 release notes

The following release notes include information for Amazon EMR release 5.28.1. Changes are relative to 5.28.0.

Initial release date: Jan 10, 2020

Changes, enhancements, and resolved issues

- Spark

  - Fixed Spark compatibility issues.
- CloudWatch Metrics
  - Fixed Amazon CloudWatch Metrics publishing on an EMR cluster with multiple primary nodes.
- Disabled log message
  - Disabled false log message, "...using old version (<4.5.8) of Apache http client."

Known issues

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as `hadoop` user to the lead primary node of the EMR cluster with multiple primary nodes.
  - Run the following command to renew Kerberos ticket for `hadoop` user.

    ```
    kinit -kt <keytab_file> <principal>
    ```

    Typically, the keytab file is located at `/etc/hadoop.keytab` and the principal is in the form of `hadoop/<hostname>@<REALM>`.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

5.28.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.2.6	Amazon SageMaker Spark SDK
emr-ddb	4.12.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.11.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.13.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.3.0	EMR S3Select Connector
emrfs	2.37.0	Amazon S3 connector for Hadoop ecosystem applicati ons.

Component	Version	Description
flink-client	1.9.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.5-amzn-5	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-5	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.5-amzn-5	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.5-amzn-5	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	2.8.5-amzn-5	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.8.5-amzn-5	HTTP endpoint for HDFS operations.

Component	Version	Description
hadoop-kms-server	2.8.5-amzn-5	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.5-amzn-5	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-5	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.5-amzn-5	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.5-amzn-5	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.10	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.10	Service for serving one or more HBase regions.
hbase-client	1.4.10	HBase command-line client.
hbase-rest-server	1.4.10	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.10	Service providing a Thrift endpoint to HBase.

Component	Version	Description
hcatalog-client	2.3.6-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.6-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.6-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.6-amzn-0	Hive command line client.
hive-hbase	2.3.6-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.6-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.6-amzn-0	Service for accepting Hive queries as web requests.
hudi	0.5.0-incubating	Incremental processing framework to power data pipline at low latency and high efficiency.
hudi-presto	0.5.0-incubating	Bundle library for running Presto with Hudi.
hue-server	4.4.0	Web application for analyzing data using Hadoop ecosystem applications

Component	Version	Description
jupyterhub	1.0.0	Multi-user server for Jupyter notebooks
livy-server	0.6.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.5.1	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.1.0	Oozie command-line client.
oozie-server	5.1.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.3-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API

Component	Version	Description
presto-coordinator	0.227	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.227	Service for executing pieces of a query.
presto-client	0.227	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.4	Spark command-line clients.
spark-history-server	2.4.4	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.4	In-memory execution engine for YARN.
spark-yarn-slave	2.4.4	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.

Component	Version	Description
tensorflow	1.14.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.2	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

5.28.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.28.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.

Classifications	Description
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.

Classifications	Description
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.

Classifications	Description
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.

Classifications	Description
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.

Classifications	Description
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.

Classifications	Description
ranger-kms-dbks-site	Change values in dbks-site.xml file of Ranger KMS.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.
ranger-kms-env	Change values in the Ranger KMS environme nt.
ranger-kms-log4j	Change values in kms-log4j.properties file of Ranger KMS.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.
recordserver-env	Change values in the EMR RecordServer environment.
recordserver-conf	Change values in EMR RecordServer's erver.pro perties file.
recordserver-log4j	Change values in EMR RecordServer's log4j.pro perties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.

Classifications	Description
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.28.0

5.28.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hudi, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases

- [Application versions in Amazon EMR 4.x releases]

Application version information

	emr-5.28.0	emr-5.27.1	emr-5.27.0	emr-5.26.0
AWS SDK for Java	1.11.659	1.11.615	1.11.615	1.11.595
Python	2.7, 3.6	2.7, 3.6	2.7, 3.6	2.7, 3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.9.0	1.8.1	1.8.1	1.8.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.10	1.4.10	1.4.10
HCatalog	2.3.6	2.3.5	2.3.5	2.3.5
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.5	2.3.5	2.3.5
Hudi	0.5.0-incubating	-	-	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	1.0.0	1.0.0	0.9.6

	emr-5.28.0	emr-5.27.1	emr-5.27.0	emr-5.26.0
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.5.1	1.4.0	1.4.0	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.3	4.14.2	4.14.2	4.14.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.227	0.224	0.224	0.220
Spark	2.4.4	2.4.4	2.4.4	2.4.3
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.13.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

5.28.0 release notes

The following release notes include information for Amazon EMR release 5.28.0. Changes are relative to 5.27.0.

Initial release date: Nov 12, 2019

Upgrades

- Upgraded Flink to version 1.9.0

- Upgraded Hive to version 2.3.6
- Upgraded MXNet to version 1.5.1
- Upgraded Phoenix to version 4.14.3
- Upgraded Presto to version 0.227
- Upgraded Zeppelin to version 0.8.2

New features

- Apache Hudi is now available for Amazon EMR to install when you create a cluster. For more information, see Hudi.
- (Nov 25, 2019) You can now choose to run multiple steps in parallel to improve cluster utilization and save cost. You can also cancel both pending and running steps. For more information, see Work with Steps Using the AWS CLI and Console.
- (Dec 3, 2019) You can now create and run EMR clusters on AWS Outposts. AWS Outposts enables native AWS services, infrastructure, and operating models in on-premises facilities. In AWS Outposts environments, you can use the same AWS APIs, tools, and infrastructure that you use in the AWS cloud. For more information, see EMR clusters on AWS Outposts.
- (Mar 11, 2020) Beginning with Amazon EMR version 5.28.0, you can create and run Amazon EMR clusters on an AWS Local Zones subnet as a logical extension of an AWS Region that supports Local Zones. A Local Zone enables Amazon EMR features and a subset of AWS services, like compute and storage services, to be located closer to users, providing very low latency access to applications running locally. For a list of available Local Zones, see AWS Local Zones. For information about accessing available AWS Local Zones, see Regions, Availability Zones, and Local Zones.

  Local Zones do not currently support Amazon EMR Notebooks and do not support connections directly to Amazon EMR using interface VPC endpoint (AWS PrivateLink).

Changes, enhancements, and resolved issues

- Expanded Application Support for High Availability Clusters
  - For more information, see Supported applications in an EMR cluster with Multiple Primary Nodes in the *Amazon EMR Management Guide*.
- Spark
  - Performance optimizations

- Hive
  - Performance optimizations
- Presto
  - Performance optimizations

Known issues

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as `hadoop` user to the lead primary node of the EMR cluster with multiple primary nodes.
  - Run the following command to renew Kerberos ticket for `hadoop` user.

    ```
    kinit -kt <keytab_file> <principal>
    ```

    Typically, the keytab file is located at `/etc/hadoop.keytab` and the principal is in the form of `hadoop/<hostname>@<REALM>`.

  > **Note**
  >
  > This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

5.28.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the

most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
aws-sagemaker-spark-sdk	1.2.6	Amazon SageMaker Spark SDK
emr-ddb	4.12.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.11.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.13.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.3.0	EMR S3Select Connector
emrfs	2.37.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.9.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati

Component	Version	Description
		ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.5-amzn-5	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-5	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.5-amzn-5	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.5-amzn-5	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	2.8.5-amzn-5	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.8.5-amzn-5	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.5-amzn-5	Cryptographic key management server based on Hadoop's KeyProvider API.

Component	Version	Description
hadoop-mapred	2.8.5-amzn-5	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-5	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.5-amzn-5	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.5-amzn-5	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.10	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.10	Service for serving one or more HBase regions.
hbase-client	1.4.10	HBase command-line client.
hbase-rest-server	1.4.10	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.10	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.6-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.

Component	Version	Description
hcatalog-server	2.3.6-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.6-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.6-amzn-0	Hive command line client.
hive-hbase	2.3.6-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.6-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.6-amzn-0	Service for accepting Hive queries as web requests.
hudi	0.5.0-incubating	Incremental processing framework to power data pipline at low latency and high efficiency.
hudi-presto	0.5.0-incubating	Bundle library for running Presto with Hudi.
hue-server	4.4.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.0.0	Multi-user server for Jupyter notebooks

Component	Version	Description
livy-server	0.6.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.5.1	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.1.0	Oozie command-line client.
oozie-server	5.1.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.3-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API

Component	Version	Description
presto-coordinator	0.227	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.227	Service for executing pieces of a query.
presto-client	0.227	Presto command-line client which is installed on an HA cluster's stand-by masters where Presto server is not started.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.4	Spark command-line clients.
spark-history-server	2.4.4	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.4	In-memory execution engine for YARN.
spark-yarn-slave	2.4.4	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.

Component	Version	Description
tensorflow	1.14.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.2	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

5.28.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.28.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.

Classifications	Description
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.

Classifications	Description
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.

Classifications	Description
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environment.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhub_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.

Classifications	Description
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.

Classifications	Description
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.

Classifications	Description
ranger-kms-dbks-site	Change values in dbks-site.xml file of Ranger KMS.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.
ranger-kms-env	Change values in the Ranger KMS environme nt.
ranger-kms-log4j	Change values in kms-log4j.properties file of Ranger KMS.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.
recordserver-env	Change values in the EMR RecordServer environment.
recordserver-conf	Change values in EMR RecordServer's erver.pro perties file.
recordserver-log4j	Change values in EMR RecordServer's log4j.pro perties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.

Classifications	Description
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.27.1

5.27.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases

- [Application versions in Amazon EMR 4.x releases]

Application version information

	emr-5.27.1	emr-5.27.0	emr-5.26.0	emr-5.25.0
AWS SDK for Java	1.11.615	1.11.615	1.11.595	1.11.566
Python	2.7, 3.6	2.7, 3.6	2.7, 3.6	2.7, 3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.8.1	1.8.1	1.8.0	1.8.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.10	1.4.10	1.4.9
HCatalog	2.3.5	2.3.5	2.3.5	2.3.5
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.5	2.3.5	2.3.5	2.3.5
Hudi	-	-	-	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	1.0.0	0.9.6	0.9.6

	emr-5.27.1	emr-5.27.0	emr-5.26.0	emr-5.25.0
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.4.0	1.4.0	1.4.0	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.2	4.14.2	4.14.2	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.224	0.224	0.220	0.220
Spark	2.4.4	2.4.4	2.4.3	2.4.3
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.13.1	1.13.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

5.27.1 release notes

This is a patch release. All applications and components are the same as the previous Amazon EMR release.

Instance Metadata Service (IMDS) V2 support status: Amazon EMR 5.23.1, 5.27.1 and 5.32 or later components use IMDSv2 for all IMDS calls. For IMDS calls in your application code, you can use both IMDSv1 and IMDSv2, or configure the IMDS to use only IMDSv2 for added security. For other 5.x EMR releases, disabling IMDSv1 causes cluster startup failure.

5.27.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.2.4	Amazon SageMaker Spark SDK
emr-ddb	4.12.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.11.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.13.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.3.0	EMR S3Select Connector
emrfs	2.36.0	Amazon S3 connector for Hadoop ecosystem applicati ons.

Component	Version	Description
flink-client	1.8.1	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.5-amzn-4	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.5-amzn-4	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.8.5-amzn-4	HTTP endpoint for HDFS operations.

Component	Version	Description
hadoop-kms-server	2.8.5-amzn-4	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.5-amzn-4	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.5-amzn-4	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.10	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.10	Service for serving one or more HBase regions.
hbase-client	1.4.10	HBase command-line client.
hbase-rest-server	1.4.10	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.10	Service providing a Thrift endpoint to HBase.

Component	Version	Description
hcatalog-client	2.3.5-amzn-1	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.5-amzn-1	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.5-amzn-1	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.5-amzn-1	Hive command line client.
hive-hbase	2.3.5-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.5-amzn-1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.5-amzn-1	Service for accepting Hive queries as web requests.
hue-server	4.4.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.0.0	Multi-user server for Jupyter notebooks
livy-server	0.6.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server

Component	Version	Description
mahout-client	0.13.0	Library for machine learning.
mxnet	1.4.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.1.0	Oozie command-line client.
oozie-server	5.1.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.2-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.2-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.224	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.224	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.

Component	Version	Description
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.4	Spark command-line clients.
spark-history-server	2.4.4	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.4	In-memory execution engine for YARN.
spark-yarn-slave	2.4.4	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	1.14.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.1	Web-based notebook that enables interactive data analytics.

Component	Version	Description
zookeeper-server	3.4.14	Centralized service for maintaining configuration information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

5.27.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.27.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.

Classifications	Description
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.

Classifications	Description
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.

Classifications	Description
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.

Classifications	Description
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.

Classifications	Description
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.
ranger-kms-dbks-site	Change values in dbks-site.xml file of Ranger KMS.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.
ranger-kms-env	Change values in the Ranger KMS environme nt.
ranger-kms-log4j	Change values in kms-log4j.properties file of Ranger KMS.

Classifications	Description
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.
recordserver-env	Change values in the EMR RecordServer environment.
recordserver-conf	Change values in EMR RecordServer's erver.pro perties file.
recordserver-log4j	Change values in EMR RecordServer's log4j.pro perties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.

Classifications	Description
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.27.0

5.27.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.27.0	emr-5.26.0	emr-5.25.0	emr-5.24.1
AWS SDK for Java	1.11.615	1.11.595	1.11.566	1.11.546
Python	2.7, 3.6	2.7, 3.6	2.7, 3.6	2.7, 3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12

	emr-5.27.0	emr-5.26.0	emr-5.25.0	emr-5.24.1
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.8.1	1.8.0	1.8.0	1.8.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.10	1.4.9	1.4.9
HCatalog	2.3.5	2.3.5	2.3.5	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.5	2.3.5	2.3.5	2.3.4
Hudi	-	-	-	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	0.9.6	0.9.6	0.9.6
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.4.0	1.4.0	1.4.0	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.2	4.14.2	4.14.1	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-5.27.0	emr-5.26.0	emr-5.25.0	emr-5.24.1
Presto	0.224	0.220	0.220	0.219
Spark	2.4.4	2.4.3	2.4.3	2.4.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.13.1	1.13.1	1.12.0
Tez	0.9.2	0.9.2	0.9.2	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.13

5.27.0 release notes

The following release notes include information for Amazon EMR release 5.27.0. Changes are relative to 5.26.0.

Initial release date: Sep 23, 2019

Upgrades

- AWS SDK for Java 1.11.615
- Flink 1.8.1
- JupyterHub 1.0.0
- Spark 2.4.4
- Tensorflow 1.14.0
- Connectors and drivers:
  - DynamoDB Connector 4.12.0

New features

- (Oct 24, 2019) The following New features in EMR notebooks are available with all Amazon EMR releases.
  - You can now associate Git repositories with EMR notebooks to store your notebooks in a version controlled environment. You can share code with peers and reuse existing Jupyter notebooks through remote Git repositories. For more information, see Associate Git Repositories with Amazon EMR Notebooks in the *Amazon EMR Management Guide*.
  - The nbdime utility is now available in EMR notebooks to simplify comparing and merging notebooks.
  - EMR notebooks now support JupyterLab. JupyterLab is a web-based interactive development environment fully compatible with Jupyter notebooks. You can now choose to open your notebook in either JupyterLab or Jupyter notebook editor.
- (Oct 30, 2019) With Amazon EMR versions 5.25.0 and later, you can connect to Spark history server UI from the cluster **Summary** page or the **Application history** tab in the console. Instead of setting up a web proxy through an SSH connection, you can quickly access the Spark history server UI to view application metrics and access relevant log files for active and terminated clusters. For more information, see Off-cluster access to persistent application user interfaces in the *Amazon EMR Management Guide*.

Changes, enhancements, and resolved issues

- Amazon EMR cluster with multiple primary nodes
  - You can install and run Flink on an Amazon EMR cluster with multiple primary nodes. For more information, see Supported applications and features.
  - You can configure HDFS transparent encryption on an Amazon EMR cluster with multiple primary nodes. For more information, see HDFS Transparent Encryption on EMR clusters with Multiple Primary Nodes.
  - You can now modify the configuration of applications running on an Amazon EMR cluster with multiple primary nodes. For more information, see Supplying a Configuration for an Instance Group in a Running Cluster.
- Amazon EMR-DynamoDB Connector
  - Amazon EMR-DynamoDB Connector now supports the following DynamoDB data types: boolean, list, map, item, null. For more information, see Set Up a Hive Table to Run Hive Commands.

Known issues

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as `hadoop` user to the lead primary node of the EMR cluster with multiple primary nodes.
  - Run the following command to renew Kerberos ticket for `hadoop` user.

    ```
    kinit -kt <keytab_file> <principal>
    ```

    Typically, the keytab file is located at `/etc/hadoop.keytab` and the principal is in the form of `hadoop/<hostname>@<REALM>`.

  > **Note**
  >
  > This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

5.27.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2

has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.2.4	Amazon SageMaker Spark SDK
emr-ddb	4.12.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.11.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.13.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.3.0	EMR S3Select Connector
emrfs	2.36.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.8.1	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.

Component	Version	Description
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.5-amzn-4	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.5-amzn-4	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.8.5-amzn-4	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.5-amzn-4	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.5-amzn-4	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARN service for managing containers on an individual node.

Component	Version	Description
hadoop-yarn-resour cemanager	2.8.5-amzn-4	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.10	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.10	Service for serving one or more HBase regions.
hbase-client	1.4.10	HBase command-line client.
hbase-rest-server	1.4.10	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.10	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.5-amzn-1	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.5-amzn-1	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.5-amzn-1	HTTP endpoint providing a REST interface to HCatalog.

Component	Version	Description
hive-client	2.3.5-amzn-1	Hive command line client.
hive-hbase	2.3.5-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.5-amzn-1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.5-amzn-1	Service for accepting Hive queries as web requests.
hue-server	4.4.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	1.0.0	Multi-user server for Jupyter notebooks
livy-server	0.6.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.4.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.1.0	Oozie command-line client.

Component	Version	Description
oozie-server	5.1.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.2-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.2-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.224	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.224	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.4	Spark command-line clients.
spark-history-server	2.4.4	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.4	In-memory execution engine for YARN.

Component	Version	Description
spark-yarn-slave	2.4.4	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	1.14.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

5.27.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see [Configure applications](#).

emr-5.27.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.

Classifications	Description
hbase-metrics	Change values in HBase's hadoop-metrics2-hbase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.

Classifications	Description
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.

Classifications	Description
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.

Classifications	Description
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.

Classifications	Description
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.
ranger-kms-dbks-site	Change values in dbks-site.xml file of Ranger KMS.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.
ranger-kms-env	Change values in the Ranger KMS environme nt.
ranger-kms-log4j	Change values in kms-log4j.properties file of Ranger KMS.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.
recordserver-env	Change values in the EMR RecordServer environment.
recordserver-conf	Change values in EMR RecordServer's erver.pro perties file.
recordserver-log4j	Change values in EMR RecordServer's log4j.pro perties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.

Classifications	Description
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.26.0

5.26.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.26.0	emr-5.25.0	emr-5.24.1	emr-5.24.0
AWS SDK for Java	1.11.595	1.11.566	1.11.546	1.11.546
Python	2.7, 3.6	2.7, 3.6	2.7, 3.6	2.7, 3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.8.0	1.8.0	1.8.0	1.8.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.9	1.4.9	1.4.9
HCatalog	2.3.5	2.3.5	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.5	2.3.5	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0

	emr-5.26.0	emr-5.25.0	emr-5.24.1	emr-5.24.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.6	0.9.6	0.9.6	0.9.6
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.4.0	1.4.0	1.4.0	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.2	4.14.1	4.14.1	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.220	0.220	0.219	0.219
Spark	2.4.3	2.4.3	2.4.2	2.4.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.13.1	1.13.1	1.12.0	1.12.0
Tez	0.9.2	0.9.2	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.13	3.4.13

5.26.0 release notes

The following release notes include information for Amazon EMR release 5.26.0. Changes are relative to 5.25.0.

Initial release date: Aug 8, 2019

Last updated date: Aug 19, 2019

Upgrades

- AWS SDK for Java 1.11.595
- HBase 1.4.10
- Phoenix 4.14.2
- Connectors and drivers:
  - DynamoDB Connector 4.11.0
  - MariaDB Connector 2.4.2
  - Amazon Redshift JDBC Driver 1.2.32.1056

New features

- (Beta) With Amazon EMR 5.26.0, you can launch a cluster that integrates with Lake Formation. This integration provides fine-grained, column-level access to databases and tables in the AWS Glue Data Catalog. It also enables federated single sign-on to EMR Notebooks or Apache Zeppelin from an enterprise identity system. For more information, see Integrating Amazon EMR with AWS Lake Formation (Beta).
- (Aug 19, 2019) Amazon EMR block public access is now available with all Amazon EMR releases that support security groups. Block public access is an account-wide setting applied to each AWS Region. Block public access prevents a cluster from launching when any security group associated with the cluster has a rule that allows inbound traffic from IPv4 0.0.0.0/0 or IPv6 ::/0 (public access) on a port, unless a port is specified as an exception. Port 22 is an exception by default. For more information, see Using Amazon EMR Block Public Access in the *Amazon EMR Management Guide*.

Changes, enhancements, and resolved issues

- EMR Notebooks

  - With EMR 5.26.0 and later, EMR Notebooks supports notebook-scoped Python libraries in addition to the default Python libraries. You can install notebook-scoped libraries from within the notebook editor without having to re-create a cluster or re-attach a notebook to a cluster. Notebook-scoped libraries are created in a Python virtual environment, so they apply only to the current notebook session. This allows you to isolate notebook dependencies. For more information, see Using Notebook Scoped Libraries in the *Amazon EMR Management Guide*.

- EMRFS
  - You can enable an ETag verification feature (Beta) by setting `fs.s3.consistent.metadata.etag.verification.enabled` to `true`. With this feature, EMRFS uses Amazon S3 ETags to verify that objects being read are the latest available version. This feature is helpful for read-after-update use cases in which files on Amazon S3 are overwritten while retaining the same name. This ETag verification capability currently does not work with S3 Select. For more information, see Configure Consistent View.

- Spark
  - The following optimizations are now enabled by default: dynamic partition pruning, DISTINCT before INTERSECT, improvements in SQL plan statistics inference for JOIN followed by DISTINCT queries, flattening scalar subqueries, optimized join reorder, and bloom filter join. For more information, see Optimizing Spark Performance.
  - Improved whole stage code generation for Sort Merge Join.
  - Improved query fragment and subquery reuse.
  - Improvements to pre-allocate executors on Spark start up.
  - Bloom filter joins are no longer applied when the smaller side of the join includes a broadcast hint.

- Tez
  - Resolved an issue with Tez. Tez UI now works on an Amazon EMR cluster with multiple primary nodes.

Known issues

- The improved whole stage code generation capabilities for Sort Merge Join can increase memory pressure when enabled. This optimization improves performance, but may result in job retries or failures if the `spark.yarn.executor.memoryOverheadFactor` is not tuned to provide enough memory. To disable this feature, set `spark.sql.sortMergeJoinExec.extendedCodegen.enabled` to false.

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as `hadoop` user to the lead primary node of the EMR cluster with multiple primary nodes.
  - Run the following command to renew Kerberos ticket for `hadoop` user.

    ```
    kinit -kt <keytab_file> <principal>
    ```

    Typically, the keytab file is located at `/etc/hadoop.keytab` and the principal is in the form of `hadoop/<hostname>@<REALM>`.

  Note

  This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

5.26.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.2.4	Amazon SageMaker Spark SDK
emr-ddb	4.11.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.10.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.12.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.3.0	EMR S3Select Connector
emrfs	2.35.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.8.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.

Component	Version	Description
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.5-amzn-4	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.5-amzn-4	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.8.5-amzn-4	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.5-amzn-4	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.5-amzn-4	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARN service for managing containers on an individual node.

Component	Version	Description
hadoop-yarn-resour cemanager	2.8.5-amzn-4	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.10	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.10	Service for serving one or more HBase regions.
hbase-client	1.4.10	HBase command-line client.
hbase-rest-server	1.4.10	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.10	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.5-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.5-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.5-amzn-0	HTTP endpoint providing a REST interface to HCatalog.

Component	Version	Description
hive-client	2.3.5-amzn-0	Hive command line client.
hive-hbase	2.3.5-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.5-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.5-amzn-0	Service for accepting Hive queries as web requests.
hue-server	4.4.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.9.6	Multi-user server for Jupyter notebooks
livy-server	0.6.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.4.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.1.0	Oozie command-line client.

Component	Version	Description
oozie-server	5.1.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.2-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.2-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.220	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.220	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.3	Spark command-line clients.
spark-history-server	2.4.3	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.3	In-memory execution engine for YARN.

Component	Version	Description
spark-yarn-slave	2.4.3	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	1.13.1	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

5.26.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.26.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.

Classifications	Description
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.

Classifications	Description
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.

Classifications	Description
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.

Classifications	Description
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.

Classifications	Description
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.
recordserver-env	Change values in the EMR RecordServer environment.
recordserver-conf	Change values in EMR RecordServer's erver.pro perties file.
recordserver-log4j	Change values in EMR RecordServer's log4j.pro perties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.

Classifications	Description
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.25.0

5.25.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.25.0	emr-5.24.1	emr-5.24.0	emr-5.23.1
AWS SDK for Java	1.11.566	1.11.546	1.11.546	1.11.519

	emr-5.25.0	emr-5.24.1	emr-5.24.0	emr-5.23.1
Python	2.7, 3.6	2.7, 3.6	2.7, 3.6	2.7, 3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.8.0	1.8.0	1.8.0	1.7.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.9	1.4.9	1.4.9
HCatalog	2.3.5	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.5	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.4.0	4.4.0	4.4.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.6	0.9.6	0.9.6	0.9.4
Livy	0.6.0	0.6.0	0.6.0	0.5.0
MXNet	1.4.0	1.4.0	1.4.0	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0

	emr-5.25.0	emr-5.24.1	emr-5.24.0	emr-5.23.1
Phoenix	4.14.1	4.14.1	4.14.1	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.220	0.219	0.219	0.215
Spark	2.4.3	2.4.2	2.4.2	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.13.1	1.12.0	1.12.0	1.12.0
Tez	0.9.2	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.13	3.4.13	3.4.13

5.25.0 release notes

The following release notes include information for Amazon EMR release 5.25.0. Changes are relative to 5.24.1.

Initial release date: July 17, 2019

Last updated date: Oct 30, 2019

Amazon EMR 5.25.0

Upgrades

- AWS SDK for Java 1.11.566
- Hive 2.3.5
- Presto 0.220
- Spark 2.4.3

- TensorFlow 1.13.1
- Tez 0.9.2
- Zookeeper 3.4.14

New features

- (Oct 30, 2019) Beginning with Amazon EMR version 5.25.0, you can connect to Spark history server UI from the cluster **Summary** page or the **Application history** tab in the console. Instead of setting up a web proxy through an SSH connection, you can quickly access the Spark history server UI to view application metrics and access relevant log files for active and terminated clusters. For more information, see Off-cluster access to persistent application user interfaces in the *Amazon EMR Management Guide*.

Changes, enhancements, and resolved issues

- Spark
  - Improved the performance of some joins by using Bloom filters to pre-filter inputs. The optimization is disabled by default and can be enabled by setting the Spark configuration parameter `spark.sql.bloomFilterJoin.enabled` to `true`.
  - Improved the performance of grouping by string type columns.
  - Improved the default Spark executor memory and cores configuration of R4 instance types for clusters without HBase installed.
  - Resolved a previous issue with the dynamic partition pruning feature where the pruned table has to be on the left side of the join.
  - Improved DISTINCT before INTERSECT optimization to apply to additional cases involving aliases.
  - Improved SQL plan statistics inference for JOIN followed by DISTINCT queries. This improvement is disabled by default and can be enabled by setting the Spark configuration parameter `spark.sql.statsImprovements.enabled` to `true`. This optimization is required by the Distinct before Intersect feature and will be enabled automatically when `spark.sql.optimizer.distinctBeforeIntersect.enabled` is set to `true`.
  - Optimized join order based on table size and filters. This optimization is disabled by default and can be enabled by setting the Spark configuration parameter `spark.sql.optimizer.sizeBasedJoinReorder.enabled` to `true`.

For more information, see Optimizing Spark Performance.

- EMRFS
  - The EMRFS setting, `fs.s3.buckets.create.enabled`, is now disabled by default. With testing, we found that disabling this setting improves performance and prevents unintentional creation of S3 buckets. If your application relies on this functionality, you can enable it by setting the property `fs.s3.buckets.create.enabled` to `true` in the `emrfs-site` configuration classification. For information, see Supplying a Configuration when Creating a Cluster.
- Local Disk Encryption and S3 Encryption Improvements in Security Configurations (August 5, 2019)
  - Separated Amazon S3 encryption settings from local disk encryption settings in security configuration setup.
  - Added an option to enable EBS encryption with release 5.24.0 and later. Selecting this option encrypts the root device volume in addition to storage volumes. Previous versions required using a custom AMI to encrypt the root device volume.
  - For more information, see Encryption Options in the *Amazon EMR Management Guide*.

Known issues

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as `hadoop` user to the lead primary node of the EMR cluster with multiple primary nodes.
  - Run the following command to renew Kerberos ticket for `hadoop` user.

  ```
  kinit -kt <keytab_file> <principal>
  ```

Typically, the keytab file is located at /etc/hadoop.keytab and the principal is in the form of hadoop/<hostname>@<REALM>.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

5.25.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with emr or aws. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
aws-sagemaker-spark-sdk	1.2.4	Amazon SageMaker Spark SDK
emr-ddb	4.10.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.9.0	Extra convenience libraries for the Hadoop ecosystem.

Component	Version	Description
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.11.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.3.0	EMR S3Select Connector
emrfs	2.34.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.8.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.5-amzn-4	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFS node-level service for storing blocks.

Component	Version	Description
hadoop-hdfs-library	2.8.5-amzn-4	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.8.5-amzn-4	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.5-amzn-4	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.5-amzn-4	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.5-amzn-4	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Service for retrieving current and historical information for YARN applications.

Component	Version	Description
hbase-hmaster	1.4.9	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.9	Service for serving one or more HBase regions.
hbase-client	1.4.9	HBase command-line client.
hbase-rest-server	1.4.9	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.9	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.5-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.5-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.5-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.5-amzn-0	Hive command line client.
hive-hbase	2.3.5-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.5-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.

Component	Version	Description
hive-server2	2.3.5-amzn-0	Service for accepting Hive queries as web requests.
hue-server	4.4.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.9.6	Multi-user server for Jupyter notebooks
livy-server	0.6.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.4.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.1.0	Oozie command-line client.
oozie-server	5.1.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.1-HBase-1.4	The phoenix libraries for server and client

Component	Version	Description
phoenix-query-server	4.14.1-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.220	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.220	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.3	Spark command-line clients.
spark-history-server	2.4.3	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.3	In-memory execution engine for YARN.
spark-yarn-slave	2.4.3	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.

Component	Version	Description
tensorflow	1.13.1	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.2	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.14	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.14	ZooKeeper command line client.

5.25.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.25.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.

Classifications	Description
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.

Classifications	Description
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.

Classifications	Description
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environment.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhub_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.

Classifications	Description
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.

Classifications	Description
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.

Classifications	Description
recordserver-env	Change values in the EMR RecordServer environment.
recordserver-conf	Change values in EMR RecordServer's erver.pro perties file.
recordserver-log4j	Change values in EMR RecordServer's log4j.pro perties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.

Classifications	Description
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.24.1

5.24.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.24.1	emr-5.24.0	emr-5.23.1	emr-5.23.0
AWS SDK for Java	1.11.546	1.11.546	1.11.519	1.11.519
Python	2.7, 3.6	2.7, 3.6	2.7, 3.6	2.7, 3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12

	emr-5.24.1	emr-5.24.0	emr-5.23.1	emr-5.23.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.8.0	1.8.0	1.7.1	1.7.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.9	1.4.9	1.4.9
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.4.0	4.4.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.6	0.9.6	0.9.4	0.9.4
Livy	0.6.0	0.6.0	0.5.0	0.5.0
MXNet	1.4.0	1.4.0	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.1	4.14.1	4.14.1	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-5.24.1	emr-5.24.0	emr-5.23.1	emr-5.23.0
Presto	0.219	0.219	0.215	0.215
Spark	2.4.2	2.4.2	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

5.24.1 release notes

The following release notes include information for Amazon EMR release 5.24.1. Changes are relative to 5.24.0.

Initial release date: June 26, 2019

Changes, enhancements, and resolved issues

- Updated the default Amazon Linux AMI for Amazon EMR to include important Linux kernel security updates, including the TCP SACK Denial of Service Issue (AWS-2019-005).

Known issues

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem

impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

Workaround:

- SSH as hadoop user to the lead primary node of the EMR cluster with multiple primary nodes.
- Run the following command to renew Kerberos ticket for hadoop user.

```
kinit -kt <keytab_file> <principal>
```

Typically, the keytab file is located at /etc/hadoop.keytab and the principal is in the form of hadoop/<hostname>@<REALM>.

> **Note**
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

5.24.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with emr or aws. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
aws-sagemaker-spark-sdk	1.2.1	Amazon SageMaker Spark SDK

Component	Version	Description
emr-ddb	4.9.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.8.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.11.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.3.0	EMR S3Select Connector
emrfs	2.33.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.8.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.

Component	Version	Description
hadoop-client	2.8.5-amzn-4	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.5-amzn-4	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.8.5-amzn-4	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.5-amzn-4	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.5-amzn-4	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARN service for managing containers on an individual node.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARN service for allocating and managing cluster resources and distributed applications.

Component	Version	Description
hadoop-yarn-timeline-server	2.8.5-amzn-4	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.9	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.9	Service for serving one or more HBase regions.
hbase-client	1.4.9	HBase command-line client.
hbase-rest-server	1.4.9	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.9	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.4-amzn-2	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.4-amzn-2	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.4-amzn-2	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.4-amzn-2	Hive command line client.
hive-hbase	2.3.4-amzn-2	Hive-hbase client.

Component	Version	Description
hive-metastore-server	2.3.4-amzn-2	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.4-amzn-2	Service for accepting Hive queries as web requests.
hue-server	4.4.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.9.6	Multi-user server for Jupyter notebooks
livy-server	0.6.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.4.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.1.0	Oozie command-line client.
oozie-server	5.1.0	Service for accepting Oozie workflow requests.

Component	Version	Description
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.1-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.1-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.219	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.219	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.2	Spark command-line clients.
spark-history-server	2.4.2	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.2	In-memory execution engine for YARN.
spark-yarn-slave	2.4.2	Apache Spark libraries needed by YARN slaves.

Component	Version	Description
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	1.12.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.1	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.13	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.13	ZooKeeper command line client.

5.24.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.24.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.

Classifications	Description
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.

Classifications	Description
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.

Classifications	Description
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.

Classifications	Description
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.

Classifications	Description
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.

Classifications	Description
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.24.0

5.24.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.24.0	emr-5.23.1	emr-5.23.0	emr-5.22.0
AWS SDK for Java	1.11.546	1.11.519	1.11.519	1.11.510
Python	2.7, 3.6	2.7, 3.6	2.7, 3.6	2.7, 3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-

	emr-5.24.0	emr-5.23.1	emr-5.23.0	emr-5.22.0
Delta	-	-	-	-
Flink	1.8.0	1.7.1	1.7.1	1.7.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.9	1.4.9	1.4.9
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.4.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.6	0.9.4	0.9.4	0.9.4
Livy	0.6.0	0.5.0	0.5.0	0.5.0
MXNet	1.4.0	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.1	4.14.1	4.14.1	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.219	0.215	0.215	0.215
Spark	2.4.2	2.4.0	2.4.0	2.4.0

	emr-5.24.0	emr-5.23.1	emr-5.23.0	emr-5.22.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

5.24.0 release notes

The following release notes include information for Amazon EMR release 5.24.0. Changes are relative to 5.23.0.

Initial release date: June 11, 2019

Last updated date: August 5, 2019

Upgrades

- Flink 1.8.0
- Hue 4.4.0
- JupyterHub 0.9.6
- Livy 0.6.0
- MxNet 1.4.0
- Presto 0.219
- Spark 2.4.2
- AWS SDK for Java 1.11.546
- Connectors and drivers:
  - DynamoDB Connector 4.9.0
  - MariaDB Connector 2.4.1

- Amazon Redshift JDBC Driver 1.2.27.1051

Changes, enhancements, and resolved issues

- Spark
  - Added optimization to dynamically prune partitions. The optimization is disabled by default. To enable it, set the Spark configuration parameter `spark.sql.dynamicPartitionPruning.enabled` to `true`.
  - Improved performance of `INTERSECT` queries. This optimization is disabled by default. To enable it, set the Spark configuration parameter `spark.sql.optimizer.distinctBeforeIntersect.enabled` to `true`.
  - Added optimization to flatten scalar subqueries with aggregates that use the same relation. The optimization is disabled by default. To enable it, set the Spark configuration parameter `spark.sql.optimizer.flattenScalarSubqueriesWithAggregates.enabled` to `true`.
  - Improved whole stage code generation.

  For more information, see Optimizing Spark Performance.
- Local Disk Encryption and S3 Encryption Improvements in Security Configurations (August 5, 2019)
  - Separated Amazon S3 encryption settings from local disk encryption settings in security configuration setup.
  - Added an option to enable EBS encryption. Selecting this option encrypts the root device volume in addition to storage volumes. Previous versions required using a custom AMI to encrypt the root device volume.
  - For more information, see Encryption Options in the *Amazon EMR Management Guide*.

Known issues

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem

impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

Workaround:

- SSH as hadoop user to the lead primary node of the EMR cluster with multiple primary nodes.
- Run the following command to renew Kerberos ticket for hadoop user.

```
kinit -kt <keytab_file> <principal>
```

Typically, the keytab file is located at /etc/hadoop.keytab and the principal is in the form of hadoop/<hostname>@<REALM>.

> **Note**
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

5.24.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with emr or aws. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
aws-sagemaker-spark-sdk	1.2.1	Amazon SageMaker Spark SDK

Component	Version	Description
emr-ddb	4.9.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.8.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.11.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.3.0	EMR S3Select Connector
emrfs	2.33.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.8.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.

Component	Version	Description
hadoop-client	2.8.5-amzn-4	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.5-amzn-4	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.8.5-amzn-4	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.5-amzn-4	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.5-amzn-4	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.5-amzn-4	YARN service for allocatin g and managing cluster resources and distributed applications.

Component	Version	Description
hadoop-yarn-timeline-server	2.8.5-amzn-4	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.9	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.9	Service for serving one or more HBase regions.
hbase-client	1.4.9	HBase command-line client.
hbase-rest-server	1.4.9	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.9	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.4-amzn-2	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.4-amzn-2	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.4-amzn-2	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.4-amzn-2	Hive command line client.
hive-hbase	2.3.4-amzn-2	Hive-hbase client.

Component	Version	Description
hive-metastore-server	2.3.4-amzn-2	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.4-amzn-2	Service for accepting Hive queries as web requests.
hue-server	4.4.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.9.6	Multi-user server for Jupyter notebooks
livy-server	0.6.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.4.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.1.0	Oozie command-line client.
oozie-server	5.1.0	Service for accepting Oozie workflow requests.

Component	Version	Description
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.1-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.1-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.219	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.219	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.2	Spark command-line clients.
spark-history-server	2.4.2	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.2	In-memory execution engine for YARN.
spark-yarn-slave	2.4.2	Apache Spark libraries needed by YARN slaves.

Component	Version	Description
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	1.12.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.1	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.13	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.13	ZooKeeper command line client.

5.24.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.24.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.

Classifications	Description
hbase-metrics	Change values in HBase's hadoop-metrics2-hbase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.
hive-parquet-logging	Change values in Hive's parquet-logging.properties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.properties file.

Classifications	Description
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.

Classifications	Description
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.

Classifications	Description
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.

Classifications	Description
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.

Classifications	Description
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.23.1

5.23.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.23.1	emr-5.23.0	emr-5.22.0	emr-5.21.2
AWS SDK for Java	1.11.519	1.11.519	1.11.510	1.11.479
Python	2.7, 3.6	2.7, 3.6	2.7, 3.6	2.7, 3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-

	emr-5.23.1	emr-5.23.0	emr-5.22.0	emr-5.21.2
Delta	-	-	-	-
Flink	1.7.1	1.7.1	1.7.1	1.7.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.9	1.4.9	1.4.8
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEn terpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.0.0
Phoenix	4.14.1	4.14.1	4.14.1	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.215	0.215	0.215
Spark	2.4.0	2.4.0	2.4.0	2.4.0

	emr-5.23.1	emr-5.23.0	emr-5.22.0	emr-5.21.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

5.23.1 release notes

This is a patch release. All applications and components are the same as the previous Amazon EMR release.

Instance Metadata Service (IMDS) V2 support status: Amazon EMR 5.23.1, 5.27.1 and 5.32 or later components use IMDSv2 for all IMDS calls. For IMDS calls in your application code, you can use both IMDSv1 and IMDSv2, or configure the IMDS to use only IMDSv2 for added security. For other 5.x EMR releases, disabling IMDSv1 causes cluster startup failure.

5.23.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.2.1	Amazon SageMaker Spark SDK
emr-ddb	4.8.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.7.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.11.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.2.0	EMR S3Select Connector
emrfs	2.32.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.7.1	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.

Component	Version	Description
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.5-amzn-3	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-3	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.5-amzn-3	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.5-amzn-3	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	2.8.5-amzn-3	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.8.5-amzn-3	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.5-amzn-3	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.5-amzn-3	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-3	YARN service for managing containers on an individual node.

Component	Version	Description
hadoop-yarn-resour cemanager	2.8.5-amzn-3	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.5-amzn-3	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.9	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.9	Service for serving one or more HBase regions.
hbase-client	1.4.9	HBase command-line client.
hbase-rest-server	1.4.9	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.9	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.4-amzn-1	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.4-amzn-1	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.4-amzn-1	HTTP endpoint providing a REST interface to HCatalog.

Component	Version	Description
hive-client	2.3.4-amzn-1	Hive command line client.
hive-hbase	2.3.4-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.4-amzn-1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.4-amzn-1	Service for accepting Hive queries as web requests.
hue-server	4.3.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.9.4	Multi-user server for Jupyter notebooks
livy-server	0.5.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.3.1	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.1.0	Oozie command-line client.

Component	Version	Description
oozie-server	5.1.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.1-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.1-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.215	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.215	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.0	Spark command-line clients.
spark-history-server	2.4.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.0	In-memory execution engine for YARN.

Component	Version	Description
spark-yarn-slave	2.4.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	1.12.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.1	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.13	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.13	ZooKeeper command line client.

5.23.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.23.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.

Classifications	Description
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.

Classifications	Description
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.

Classifications	Description
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.

Classifications	Description
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.

Classifications	Description
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.

Classifications	Description
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.23.0

5.23.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.23.0	emr-5.22.0	emr-5.21.2	emr-5.21.1
AWS SDK for Java	1.11.519	1.11.510	1.11.479	1.11.479
Python	2.7, 3.6	2.7, 3.6	2.7, 3.6	2.7, 3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-

	emr-5.23.0	emr-5.22.0	emr-5.21.2	emr-5.21.1
Delta	-	-	-	-
Flink	1.7.1	1.7.1	1.7.0	1.7.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.9	1.4.8	1.4.8
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.0.0	5.0.0
Phoenix	4.14.1	4.14.1	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.215	0.215	0.215
Spark	2.4.0	2.4.0	2.4.0	2.4.0

	emr-5.23.0	emr-5.22.0	emr-5.21.2	emr-5.21.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

5.23.0 release notes

The following release notes include information for Amazon EMR release 5.23.0. Changes are relative to 5.22.0.

Initial release date: April 01, 2019

Last updated date: April 30, 2019

Upgrades

- AWS SDK for Java 1.11.519

New features

- (April 30, 2019) With Amazon EMR 5.23.0 and later, you can launch a cluster with three primary nodes to support high availability of applications like YARN Resource Manager, HDFS NameNode, Spark, Hive, and Ganglia. The primary node is no longer a potential single point of failure with this feature. If one of the primary nodes fails, Amazon EMR automatically fails over to a standby primary node and replaces the failed primary node with a new one with the same configuration and bootstrap actions. For more information, see Plan and Configure Primary Nodes.

Known issues

- Tez UI (Fixed in Amazon EMR release 5.26.0)

  Tez UI does not work on an EMR cluster with multiple primary nodes.
- Hue (Fixed in Amazon EMR release 5.24.0)
  - Hue running on Amazon EMR does not support Solr. Beginning with Amazon EMR release 5.20.0, a misconfiguration issue causes Solr to be enabled and a harmless error message to appear similar to the following:

    ```
    Solr server could not be contacted properly:
    HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
    port=1978): Max retries exceeded with url: /solr/admin/info/
    system?user.name=hue&doAs=administrator&wt=json (Caused by
    NewConnectionError(': Failed to establish a new connection: [Errno 111]
    Connection refused',))
    ```

    To prevent the Solr error message from appearing:

    1. Connect to the primary node command line using SSH.
    2. Use a text editor to open the `hue.ini` file. For example:

       ```
       sudo vim /etc/hue/conf/hue.ini
       ```
    3. Search for the term `appblacklist` and modify the line to the following:

       ```
       appblacklist = search
       ```
    4. Save your changes and restart Hue as shown in the following example:

       ```
       sudo stop hue; sudo start hue
       ```
- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

Workaround:

- SSH as hadoop user to the lead primary node of the EMR cluster with multiple primary nodes.
- Run the following command to renew Kerberos ticket for hadoop user.

```
kinit -kt <keytab_file> <principal>
```

  Typically, the keytab file is located at /etc/hadoop.keytab and the principal is in the form of hadoop/<hostname>@<REALM>.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

5.23.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with emr or aws. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
aws-sagemaker-spark-sdk	1.2.1	Amazon SageMaker Spark SDK

Component	Version	Description
emr-ddb	4.8.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.7.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.11.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.2.0	EMR S3Select Connector
emrfs	2.32.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.7.1	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.

Component	Version	Description
hadoop-client	2.8.5-amzn-3	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-3	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.5-amzn-3	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.5-amzn-3	HDFS service for tracking file names and block locations.
hadoop-hdfs-journalnode	2.8.5-amzn-3	HDFS service for managing the Hadoop filesystem journal on HA clusters.
hadoop-httpfs-server	2.8.5-amzn-3	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.5-amzn-3	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.5-amzn-3	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-3	YARN service for managing containers on an individual node.
hadoop-yarn-resourcemanager	2.8.5-amzn-3	YARN service for allocating and managing cluster resources and distributed applications.

Component	Version	Description
hadoop-yarn-timeline-server	2.8.5-amzn-3	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.9	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.9	Service for serving one or more HBase regions.
hbase-client	1.4.9	HBase command-line client.
hbase-rest-server	1.4.9	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.9	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.4-amzn-1	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.4-amzn-1	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.4-amzn-1	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.4-amzn-1	Hive command line client.
hive-hbase	2.3.4-amzn-1	Hive-hbase client.

Component	Version	Description
hive-metastore-server	2.3.4-amzn-1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.4-amzn-1	Service for accepting Hive queries as web requests.
hue-server	4.3.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.9.4	Multi-user server for Jupyter notebooks
livy-server	0.5.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.3.1	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.1.0	Oozie command-line client.
oozie-server	5.1.0	Service for accepting Oozie workflow requests.

Component	Version	Description
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.1-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.1-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.215	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.215	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.0	Spark command-line clients.
spark-history-server	2.4.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.0	In-memory execution engine for YARN.
spark-yarn-slave	2.4.0	Apache Spark libraries needed by YARN slaves.

Component	Version	Description
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	1.12.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.1	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.13	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.13	ZooKeeper command line client.

5.23.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.23.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.

Classifications	Description
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.

Classifications	Description
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.

Classifications	Description
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.

Classifications	Description
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.

Classifications	Description
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.

Classifications	Description
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.22.0

5.22.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.22.0	emr-5.21.2	emr-5.21.1	emr-5.21.0
AWS SDK for Java	1.11.510	1.11.479	1.11.479	1.11.479
Python	2.7, 3.6	2.7, 3.6	2.7, 3.6	2.7, 3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-

	emr-5.22.0	emr-5.21.2	emr-5.21.1	emr-5.21.0
Delta	-	-	-	-
Flink	1.7.1	1.7.0	1.7.0	1.7.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.8	1.4.8	1.4.8
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.1	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.215	0.215	0.215
Spark	2.4.0	2.4.0	2.4.0	2.4.0

	emr-5.22.0	emr-5.21.2	emr-5.21.1	emr-5.21.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

5.22.0 release notes

The following release notes include information for Amazon EMR release 5.22.0. Changes are relative to 5.21.0.

> **Important**
>
> Beginning with Amazon EMR release 5.22.0, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. Earlier Amazon EMR releases use AWS Signature Version 2 in some cases, unless the release notes indicate that Signature Version 4 is used exclusively. For more information, see Authenticating Requests (AWS Signature Version 4) and Authenticating Requests (AWS Signature Version 2) in the *Amazon Simple Storage Service Developer Guide*.

Initial release date: March 20, 2019

Upgrades

- Flink 1.7.1
- HBase 1.4.9
- Oozie 5.1.0
- Phoenix 4.14.1

- Zeppelin 0.8.1
- Connectors and drivers:
  - DynamoDB Connector 4.8.0
  - MariaDB Connector 2.2.6
  - Amazon Redshift JDBC Driver 1.2.20.1043

New features

- Modified the default EBS configuration for EC2 instance types with EBS-only storage. When you create a cluster using Amazon EMR release 5.22.0 and later, the default amount of EBS storage increases based on the size of the instance. In addition, we split increased storage across multiple volumes, giving increased IOPS performance. If you want to use a different EBS instance storage configuration, you can specify it when you create an EMR cluster or add nodes to an existing cluster. For more information about the amount of storage and number of volumes allocated by default for each instance type, see Default EBS Storage for Instances in the *Amazon EMR Management Guide*.

Changes, enhancements, and resolved issues

- Spark
  - Introduced a new configuration property for Spark on YARN, `spark.yarn.executor.memoryOverheadFactor`. The value of this property is a scale factor that sets the value of memory overhead to a percentage of executor memory, with a minimum of 384 MB. If memory overhead is set explicitly using `spark.yarn.executor.memoryOverhead`, this property has no effect. The default value is `0.1875`, representing 18.75%. This default for Amazon EMR leaves more space in YARN containers for executor memory overhead than the 10% default set internally by Spark. The Amazon EMR default of 18.75% empirically showed fewer memory-related failures in TPC-DS benchmarks.
  - Backported SPARK-26316 to improve performance.
- In Amazon EMR version 5.19.0, 5.20.0, and 5.21.0, YARN node labels are stored in an HDFS directory. In some situations, this leads to core node startup delays and then causes cluster time-out and launch failure. Beginning with Amazon EMR 5.22.0, this issue is resolved. YARN node labels are stored on the local disk of each cluster node, avoiding dependencies on HDFS.

Known issues

- Hue (Fixed in Amazon EMR release 5.24.0)
  - Hue running on Amazon EMR does not support Solr. Beginning with Amazon EMR release 5.20.0, a misconfiguration issue causes Solr to be enabled and a harmless error message to appear similar to the following:

    ```
    Solr server could not be contacted properly:
    HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
    port=1978): Max retries exceeded with url: /solr/admin/info/
    system?user.name=hue&doAs=administrator&wt=json (Caused by
    NewConnectionError(': Failed to establish a new connection: [Errno 111]
    Connection refused',))
    ```

    To prevent the Solr error message from appearing:

    1. Connect to the primary node command line using SSH.
    2. Use a text editor to open the `hue.ini` file. For example:

       ```
       sudo vim /etc/hue/conf/hue.ini
       ```

    3. Search for the term `appblacklist` and modify the line to the following:

       ```
       appblacklist = search
       ```

    4. Save your changes and restart Hue as shown in the following example:

       ```
       sudo stop hue; sudo start hue
       ```

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as hadoop user to the lead primary node of the EMR cluster with multiple primary nodes.

- Run the following command to renew Kerberos ticket for hadoop user.

```
kinit -kt <keytab_file> <principal>
```

Typically, the keytab file is located at /etc/hadoop.keytab and the principal is in the form of hadoop/<hostname>@<REALM>.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

5.22.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with emr or aws. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
aws-sagemaker-spark-sdk	1.2.1	Amazon SageMaker Spark SDK
emr-ddb	4.8.0	Amazon DynamoDB connector for Hadoop ecosystem applications.

Component	Version	Description
emr-goodies	2.6.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.11.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.2.0	EMR S3Select Connector
emrfs	2.31.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.7.1	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.5-amzn-2	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.

Component	Version	Description
hadoop-hdfs-datanode	2.8.5-amzn-2	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.5-amzn-2	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.5-amzn-2	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.5-amzn-2	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.5-amzn-2	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.5-amzn-2	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-2	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.5-amzn-2	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.5-amzn-2	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.9	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.

Component	Version	Description
hbase-region-server	1.4.9	Service for serving one or more HBase regions.
hbase-client	1.4.9	HBase command-line client.
hbase-rest-server	1.4.9	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.9	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.4-amzn-1	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.4-amzn-1	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.4-amzn-1	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.4-amzn-1	Hive command line client.
hive-hbase	2.3.4-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.4-amzn-1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.4-amzn-1	Service for accepting Hive queries as web requests.

Component	Version	Description
hue-server	4.3.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.9.4	Multi-user server for Jupyter notebooks
livy-server	0.5.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.3.1	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.1.0	Oozie command-line client.
oozie-server	5.1.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.1-HBase-1.4	The phoenix libraries for server and client

Component	Version	Description
phoenix-query-server	4.14.1-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.215	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.215	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.0	Spark command-line clients.
spark-history-server	2.4.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.0	In-memory execution engine for YARN.
spark-yarn-slave	2.4.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.

Component	Version	Description
tensorflow	1.12.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.1	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.13	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.13	ZooKeeper command line client.

5.22.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.22.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.

Classifications	Description
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.

Classifications	Description
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.

Classifications	Description
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.

Classifications	Description
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.

Classifications	Description
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.

Classifications	Description
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.21.2

5.21.2 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.21.2	emr-5.21.1	emr-5.21.0	emr-5.20.1
AWS SDK for Java	1.11.479	1.11.479	1.11.479	1.11.461
Python	2.7, 3.6	2.7, 3.6	2.7, 3.6	2.7, 3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.7.0	1.7.0	1.7.0	1.6.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.8	1.4.8	1.4.8

	emr-5.21.2	emr-5.21.1	emr-5.21.0	emr-5.20.1
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.215	0.215	0.214
Spark	2.4.0	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1

	emr-5.21.2	emr-5.21.1	emr-5.21.0	emr-5.20.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

5.21.2 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

> **Important**
>
> In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

5.21.2 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.2.1	Amazon SageMaker Spark SDK

Component	Version	Description
emr-ddb	4.7.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.5.1	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.11.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.2.0	EMR S3Select Connector
emrfs	2.30.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.7.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.

Component	Version	Description
hadoop-client	2.8.5-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.5-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.5-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.5-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.5-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.5-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Service for retrieving current and historical information for YARN applications.

Component	Version	Description
hbase-hmaster	1.4.8	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.8	Service for serving one or more HBase regions.
hbase-client	1.4.8	HBase command-line client.
hbase-rest-server	1.4.8	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.8	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.4-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.4-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.4-amzn-0	Hive command line client.
hive-hbase	2.3.4-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.4-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.

Component	Version	Description
hive-server2	2.3.4-amzn-0	Service for accepting Hive queries as web requests.
hue-server	4.3.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.9.4	Multi-user server for Jupyter notebooks
livy-server	0.5.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.3.1	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.0.0	Oozie command-line client.
oozie-server	5.0.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	The phoenix libraries for server and client

Component	Version	Description
phoenix-query-server	4.14.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.215	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.215	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.0	Spark command-line clients.
spark-history-server	2.4.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.0	In-memory execution engine for YARN.
spark-yarn-slave	2.4.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.

Component	Version	Description
tensorflow	1.12.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.1	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.0	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.13	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.13	ZooKeeper command line client.

5.21.2 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.21.2 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.

Classifications	Description
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.

Classifications	Description
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.

Classifications	Description
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.

Classifications	Description
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.

Classifications	Description
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.

Classifications	Description
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.21.1

5.21.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.21.1	emr-5.21.0	emr-5.20.1	emr-5.20.0
AWS SDK for Java	1.11.479	1.11.479	1.11.461	1.11.461
Python	2.7, 3.6	2.7, 3.6	2.7, 3.6	2.7, 3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.7.0	1.7.0	1.6.2	1.6.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.8	1.4.8	1.4.8

	emr-5.21.1	emr-5.21.0	emr-5.20.1	emr-5.20.0
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.215	0.214	0.214
Spark	2.4.0	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1

	emr-5.21.1	emr-5.21.0	emr-5.20.1	emr-5.20.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

5.21.1 release notes

The following release notes include information for Amazon EMR release 5.21.1. Changes are relative to 5.21.0.

Initial release date: July 18, 2019

Changes, enhancements, and resolved issues

- Updated the default Amazon Linux AMI for Amazon EMR to include important Linux kernel security updates, including the TCP SACK Denial of Service Issue (AWS-2019-005).

Known issues

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as `hadoop` user to the lead primary node of the EMR cluster with multiple primary nodes.
  - Run the following command to renew Kerberos ticket for `hadoop` user.

    ```
    kinit -kt <keytab_file> <principal>
    ```

Typically, the keytab file is located at `/etc/hadoop.keytab` and the principal is in the form of `hadoop/<hostname>@<REALM>`.

Note

This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

5.21.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.2.1	Amazon SageMaker Spark SDK
emr-ddb	4.7.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.5.1	Extra convenience libraries for the Hadoop ecosystem.

Component	Version	Description
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.11.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.2.0	EMR S3Select Connector
emrfs	2.30.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.7.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.5-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFS node-level service for storing blocks.

Component	Version	Description
hadoop-hdfs-library	2.8.5-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.5-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.5-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.5-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.5-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.8	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.8	Service for serving one or more HBase regions.

Component	Version	Description
hbase-client	1.4.8	HBase command-line client.
hbase-rest-server	1.4.8	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.8	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.4-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.4-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.4-amzn-0	Hive command line client.
hive-hbase	2.3.4-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.4-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.4-amzn-0	Service for accepting Hive queries as web requests.
hue-server	4.3.0	Web application for analyzing data using Hadoop ecosystem applications

Component	Version	Description
jupyterhub	0.9.4	Multi-user server for Jupyter notebooks
livy-server	0.5.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.3.1	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.0.0	Oozie command-line client.
oozie-server	5.0.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API

Component	Version	Description
presto-coordinator	0.215	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.215	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.0	Spark command-line clients.
spark-history-server	2.4.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.0	In-memory execution engine for YARN.
spark-yarn-slave	2.4.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	1.12.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.1	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.

Component	Version	Description
zeppelin-server	0.8.0	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.13	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.13	ZooKeeper command line client.

5.21.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.21.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.

Classifications	Description
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.

Classifications	Description
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.

Classifications	Description
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.

Classifications	Description
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authenticator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.

Classifications	Description
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file

Classifications	Description
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.21.0

5.21.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- [Application versions in Amazon EMR 7.x releases]
- [Application versions in Amazon EMR 6.x releases]
- [Application versions in Amazon EMR 5.x releases]
- [Application versions in Amazon EMR 4.x releases]

Application version information

	emr-5.21.0	emr-5.20.1	emr-5.20.0	emr-5.19.1
AWS SDK for Java	1.11.479	1.11.461	1.11.461	1.11.433
Python	2.7, 3.6	2.7, 3.6	2.7, 3.6	2.7, 3.4
Scala	2.11.12	2.11.12	2.11.12	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.7.0	1.6.2	1.6.2	1.6.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.8	1.4.8	1.4.7
HCatalog	2.3.4	2.3.4	2.3.4	2.3.3
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.3
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.2.0
Iceberg	-	-	-	-

	emr-5.21.0	emr-5.20.1	emr-5.20.0	emr-5.19.1
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.1	1.3.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.214	0.214	0.212
Spark	2.4.0	2.4.0	2.4.0	2.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.11.0
Tez	0.9.1	0.9.1	0.9.1	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

5.21.0 release notes

The following release notes include information for Amazon EMR release 5.21.0. Changes are relative to 5.20.0.

Initial release date: February 18, 2019

Last updated date: April 3, 2019

Upgrades

- Flink 1.7.0
- Presto 0.215
- AWS SDK for Java 1.11.479

New features

- (April 3, 2019) With Amazon EMR version 5.21.0 and later, you can override cluster configurations and specify additional configuration classifications for each instance group in a running cluster. You do this by using the Amazon EMR console, the AWS Command Line Interface (AWS CLI), or the AWS SDK. For more information, see [Supplying a Configuration for an Instance Group in a Running Cluster](#).

Changes, enhancements, and resolved issues

- Zeppelin
  - Backported [ZEPPELIN-3878](#).

Known issues

- Hue (Fixed in Amazon EMR release 5.24.0)
  - Hue running on Amazon EMR does not support Solr. Beginning with Amazon EMR release 5.20.0, a misconfiguration issue causes Solr to be enabled and a harmless error message to appear similar to the following:

    ```
    Solr server could not be contacted properly:
    HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
    port=1978): Max retries exceeded with url: /solr/admin/info/
    system?user.name=hue&doAs=administrator&wt=json (Caused by
    NewConnectionError(': Failed to establish a new connection: [Errno 111]
    Connection refused',))
    ```

    To prevent the Solr error message from appearing:

1. Connect to the primary node command line using SSH.
2. Use a text editor to open the `hue.ini` file. For example:

   `sudo vim /etc/hue/conf/hue.ini`
3. Search for the term `appblacklist` and modify the line to the following:

```
appblacklist = search
```

4. Save your changes and restart Hue as shown in the following example:

```
sudo stop hue; sudo start hue
```

- Tez
  - This issue was fixed in Amazon EMR 5.22.0.

    When you connect to the Tez UI at http://*MasterDNS*:8080/tez-ui through an SSH connection to the cluster primary node, the error "Adapter operation failed - Timeline server (ATS) is out of reach. Either it is down, or CORS is not enabled" appears, or tasks unexpectedly show N/A.

    This is caused by the Tez UI making requests to the YARN Timeline Server using `localhost` rather than the host name of the primary node. As a workaround, a script is available to run as a bootstrap action or step. The script updates the host name in the Tez `configs.env` file. For more information and the location of the script, see the Bootstrap Instructions.
- In Amazon EMR version 5.19.0, 5.20.0, and 5.21.0, YARN node labels are stored in an HDFS directory. In some situations, this leads to core node startup delays and then causes cluster time-out and launch failure. Beginning with Amazon EMR 5.22.0, this issue is resolved. YARN node labels are stored on the local disk of each cluster node, avoiding dependencies on HDFS.
- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:
  - SSH as `hadoop` user to the lead primary node of the EMR cluster with multiple primary nodes.

- Run the following command to renew Kerberos ticket for hadoop user.

```
kinit -kt <keytab_file> <principal>
```

Typically, the keytab file is located at /etc/hadoop.keytab and the principal is in the form of hadoop/<hostname>@<REALM>.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

5.21.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with emr or aws. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
aws-sagemaker-spark-sdk	1.2.1	Amazon SageMaker Spark SDK
emr-ddb	4.7.0	Amazon DynamoDB connector for Hadoop ecosystem applications.

Component	Version	Description
emr-goodies	2.5.1	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.11.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.2.0	EMR S3Select Connector
emrfs	2.30.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.7.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.5-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.

Component	Version	Description
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.5-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.5-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.5-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.5-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.5-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.8	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.

Component	Version	Description
hbase-region-server	1.4.8	Service for serving one or more HBase regions.
hbase-client	1.4.8	HBase command-line client.
hbase-rest-server	1.4.8	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.8	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.4-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.4-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.4-amzn-0	Hive command line client.
hive-hbase	2.3.4-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.4-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.4-amzn-0	Service for accepting Hive queries as web requests.

Component	Version	Description
hue-server	4.3.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.9.4	Multi-user server for Jupyter notebooks
livy-server	0.5.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.3.1	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.0.0	Oozie command-line client.
oozie-server	5.0.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	The phoenix libraries for server and client

Component	Version	Description
phoenix-query-server	4.14.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.215	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.215	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.0	Spark command-line clients.
spark-history-server	2.4.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.0	In-memory execution engine for YARN.
spark-yarn-slave	2.4.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.

Component	Version	Description
tensorflow	1.12.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.1	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.0	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.13	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.13	ZooKeeper command line client.

5.21.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.21.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.

Classifications	Description
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.

Classifications	Description
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.

Classifications	Description
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.

Classifications	Description
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.

Classifications	Description
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.

Classifications	Description
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.20.1

5.20.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.20.1	emr-5.20.0	emr-5.19.1	emr-5.19.0
AWS SDK for Java	1.11.461	1.11.461	1.11.433	1.11.433
Python	2.7, 3.6	2.7, 3.6	2.7, 3.4	2.7, 3.4
Scala	2.11.12	2.11.12	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.6.2	1.6.2	1.6.1	1.6.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.8	1.4.7	1.4.7

	emr-5.20.1	emr-5.20.0	emr-5.19.1	emr-5.19.0
HCatalog	2.3.4	2.3.4	2.3.3	2.3.3
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.0	1.3.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.214	0.214	0.212	0.212
Spark	2.4.0	2.4.0	2.3.2	2.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.11.0	1.11.0
Tez	0.9.1	0.9.1	0.8.4	0.8.4

	emr-5.20.1	emr-5.20.0	emr-5.19.1	emr-5.19.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

5.20.1 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

Important

In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

5.20.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.2.1	Amazon SageMaker Spark SDK

Component	Version	Description
emr-ddb	4.7.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.5.1	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.10.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.2.0	EMR S3Select Connector
emrfs	2.29.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.6.2	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.

Component	Version	Description
hadoop-client	2.8.5-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.5-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.5-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.5-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.5-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.5-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Service for retrieving current and historical information for YARN applications.

Component	Version	Description
hbase-hmaster	1.4.8	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.8	Service for serving one or more HBase regions.
hbase-client	1.4.8	HBase command-line client.
hbase-rest-server	1.4.8	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.8	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.4-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.4-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.4-amzn-0	Hive command line client.
hive-hbase	2.3.4-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.4-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.

Component	Version	Description
hive-server2	2.3.4-amzn-0	Service for accepting Hive queries as web requests.
hue-server	4.3.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.9.4	Multi-user server for Jupyter notebooks
livy-server	0.5.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.3.1	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.0.0	Oozie command-line client.
oozie-server	5.0.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	The phoenix libraries for server and client

Component	Version	Description
phoenix-query-server	4.14.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.214	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.214	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.0	Spark command-line clients.
spark-history-server	2.4.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.0	In-memory execution engine for YARN.
spark-yarn-slave	2.4.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.

Component	Version	Description
tensorflow	1.12.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.1	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.0	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.13	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.13	ZooKeeper command line client.

5.20.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.20.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.

Classifications	Description
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.

Classifications	Description
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.

Classifications	Description
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.

Classifications	Description
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.

Classifications	Description
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.

Classifications	Description
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.20.0

5.20.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.20.0	emr-5.19.1	emr-5.19.0	emr-5.18.1
AWS SDK for Java	1.11.461	1.11.433	1.11.433	1.11.393
Python	2.7, 3.6	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	2.11.12	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.6.2	1.6.1	1.6.1	1.6.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.7	1.4.7	1.4.7

	emr-5.20.0	emr-5.19.1	emr-5.19.0	emr-5.18.1
HCatalog	2.3.4	2.3.3	2.3.3	2.3.3
Hadoop	2.8.5	2.8.5	2.8.5	2.8.4
Hive	2.3.4	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.3.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.0	1.3.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.214	0.212	0.212	0.210
Spark	2.4.0	2.3.2	2.3.2	2.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.11.0	1.11.0	1.9.0
Tez	0.9.1	0.8.4	0.8.4	0.8.4

	emr-5.20.0	emr-5.19.1	emr-5.19.0	emr-5.18.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.12

5.20.0 release notes

The following release notes include information for Amazon EMR release 5.20.0. Changes are relative to 5.19.0.

Initial release date: December 18, 2018

Last updated date: January 22, 2019

Upgrades

- Flink 1.6.2
- HBase 1.4.8
- Hive 2.3.4
- Hue 4.3.0
- MXNet 1.3.1
- Presto 0.214
- Spark 2.4.0
- TensorFlow 1.12.0
- Tez 0.9.1
- AWS SDK for Java 1.11.461

New features

- (January 22, 2019) Kerberos in Amazon EMR has been improved to support authenticating principals from an external KDC. This centralizes principal management because multiple clusters can share a single, external KDC. In addition, the external KDC can have a cross-realm trust with an Active Directory domain. This allows all clusters to authenticate principals from

Active Directory. For more information, see Use Kerberos Authentication in the *Amazon EMR Management Guide*.

Changes, enhancements, and resolved issues

- Default Amazon Linux AMI for Amazon EMR
  - Python3 package was upgraded from python 3.4 to 3.6.
- The EMRFS S3-optimized committer
  - The EMRFS S3-optimized committer is now enabled by default, which improves write performance. For more information, see Use the EMRFS S3-optimized committer.
- Hive
  - Backported HIVE-16686.
- Glue with Spark and Hive
  - In EMR 5.20.0 or later, parallel partition pruning is enabled automatically for Spark and Hive when AWS Glue Data Catalog is used as the metastore. This change significantly reduces query planning time by executing multiple requests in parallel to retrieve partitions. The total number of segments that can be executed concurrently range between 1 and 10. The default value is 5, which is a recommended setting. You can change it by specifying the property `aws.glue.partition.num.segments` in `hive-site` configuration classification. If throttling occurs, you can turn off the feature by changing the value to 1. For more information, see AWS Glue Segment Structure.

Known issues

- Hue (Fixed in Amazon EMR release 5.24.0)
  - Hue running on Amazon EMR does not support Solr. Beginning with Amazon EMR release 5.20.0, a misconfiguration issue causes Solr to be enabled and a harmless error message to appear similar to the following:

    ```
    Solr server could not be contacted properly:
    HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
    port=1978): Max retries exceeded with url: /solr/admin/info/
    system?user.name=hue&doAs=administrator&wt=json (Caused by
    NewConnectionError(': Failed to establish a new connection: [Errno 111]
    Connection refused',))
    ```

To prevent the Solr error message from appearing:

1. Connect to the primary node command line using SSH.
2. Use a text editor to open the `hue.ini` file. For example:

   `sudo vim /etc/hue/conf/hue.ini`
3. Search for the term `appblacklist` and modify the line to the following:

```
appblacklist = search
```

4. Save your changes and restart Hue as shown in the following example:

```
sudo stop hue; sudo start hue
```

- Tez
  - This issue was fixed in Amazon EMR 5.22.0.

    When you connect to the Tez UI at http://*MasterDNS*:8080/tez-ui through an SSH connection to the cluster primary node, the error "Adapter operation failed - Timeline server (ATS) is out of reach. Either it is down, or CORS is not enabled" appears, or tasks unexpectedly show N/A.

    This is caused by the Tez UI making requests to the YARN Timeline Server using `localhost` rather than the host name of the primary node. As a workaround, a script is available to run as a bootstrap action or step. The script updates the host name in the Tez `configs.env` file. For more information and the location of the script, see the Bootstrap Instructions.
- In Amazon EMR version 5.19.0, 5.20.0, and 5.21.0, YARN node labels are stored in an HDFS directory. In some situations, this leads to core node startup delays and then causes cluster time-out and launch failure. Beginning with Amazon EMR 5.22.0, this issue is resolved. YARN node labels are stored on the local disk of each cluster node, avoiding dependencies on HDFS.
- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

Workaround:

- SSH as hadoop user to the lead primary node of the EMR cluster with multiple primary nodes.
- Run the following command to renew Kerberos ticket for hadoop user.

```
kinit -kt <keytab_file> <principal>
```

Typically, the keytab file is located at /etc/hadoop.keytab and the principal is in the form of hadoop/<hostname>@<REALM>.

Note

This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

5.20.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with emr or aws. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
aws-sagemaker-spark-sdk	1.2.1	Amazon SageMaker Spark SDK

Component	Version	Description
emr-ddb	4.7.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.5.1	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.10.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.2.0	EMR S3Select Connector
emrfs	2.29.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.6.2	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.

Component	Version	Description
hadoop-client	2.8.5-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.5-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.5-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.5-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.5-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.5-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Service for retrieving current and historical information for YARN applications.

Component	Version	Description
hbase-hmaster	1.4.8	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.8	Service for serving one or more HBase regions.
hbase-client	1.4.8	HBase command-line client.
hbase-rest-server	1.4.8	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.8	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.4-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.4-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.4-amzn-0	Hive command line client.
hive-hbase	2.3.4-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.4-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.

Component	Version	Description
hive-server2	2.3.4-amzn-0	Service for accepting Hive queries as web requests.
hue-server	4.3.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.9.4	Multi-user server for Jupyter notebooks
livy-server	0.5.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.3.1	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.0.0	Oozie command-line client.
oozie-server	5.0.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	The phoenix libraries for server and client

Component	Version	Description
phoenix-query-server	4.14.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.214	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.214	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.0	Spark command-line clients.
spark-history-server	2.4.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.4.0	In-memory execution engine for YARN.
spark-yarn-slave	2.4.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.

Component	Version	Description
tensorflow	1.12.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.9.1	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.0	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.13	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.13	ZooKeeper command line client.

5.20.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.20.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.

Classifications	Description
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.

Classifications	Description
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.

Classifications	Description
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.

Classifications	Description
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.

Classifications	Description
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.

Classifications	Description
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.19.1

5.19.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.19.1	emr-5.19.0	emr-5.18.1	emr-5.18.0
AWS SDK for Java	1.11.433	1.11.433	1.11.393	1.11.393
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.6.1	1.6.1	1.6.0	1.6.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.7	1.4.7	1.4.7	1.4.7

	emr-5.19.1	emr-5.19.0	emr-5.18.1	emr-5.18.0
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.5	2.8.5	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.0	1.3.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.212	0.212	0.210	0.210
Spark	2.3.2	2.3.2	2.3.2	2.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.11.0	1.11.0	1.9.0	1.9.0
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.19.1	emr-5.19.0	emr-5.18.1	emr-5.18.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.12	3.4.12

5.19.1 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

> **Important**
>
> In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

5.19.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form `CommunityVersion-amzn-EmrVersion`. The `EmrVersion` starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.2.0	Amazon SageMaker Spark SDK

Component	Version	Description
emr-ddb	4.7.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.5.1	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.10.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.1.0	EMR S3Select Connector
emrfs	2.28.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.6.1	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.

Component	Version	Description
hadoop-client	2.8.5-amzn-0	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-0	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.5-amzn-0	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.5-amzn-0	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.5-amzn-0	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.5-amzn-0	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.5-amzn-0	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-0	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.5-amzn-0	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.5-amzn-0	Service for retrieving current and historical information for YARN applications.

Component	Version	Description
hbase-hmaster	1.4.7	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.7	Service for serving one or more HBase regions.
hbase-client	1.4.7	HBase command-line client.
hbase-rest-server	1.4.7	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.7	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.3-amzn-2	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.3-amzn-2	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.3-amzn-2	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.3-amzn-2	Hive command line client.
hive-hbase	2.3.3-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-2	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.

Component	Version	Description
hive-server2	2.3.3-amzn-2	Service for accepting Hive queries as web requests.
hue-server	4.2.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.9.4	Multi-user server for Jupyter notebooks
livy-server	0.5.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.3.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.0.0	Oozie command-line client.
oozie-server	5.0.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	The phoenix libraries for server and client

Component	Version	Description
phoenix-query-server	4.14.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.212	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.212	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.2	Spark command-line clients.
spark-history-server	2.3.2	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.3.2	In-memory execution engine for YARN.
spark-yarn-slave	2.3.2	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.

Component	Version	Description
tensorflow	1.11.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.0	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.13	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.13	ZooKeeper command line client.

5.19.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.19.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.

Classifications	Description
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.

Classifications	Description
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.

Classifications	Description
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environment.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhub_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.

Classifications	Description
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.

Classifications	Description
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.

Classifications	Description
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.19.0

5.19.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.19.0	emr-5.18.1	emr-5.18.0	emr-5.17.2
AWS SDK for Java	1.11.433	1.11.393	1.11.393	1.11.336
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.6.1	1.6.0	1.6.0	1.5.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.7	1.4.7	1.4.7	1.4.6

	emr-5.19.0	emr-5.18.1	emr-5.18.0	emr-5.17.2
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.5	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.0	1.2.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.212	0.210	0.210	0.206
Spark	2.3.2	2.3.2	2.3.2	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.11.0	1.9.0	1.9.0	1.9.0
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.19.0	emr-5.18.1	emr-5.18.0	emr-5.17.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.7.3
ZooKeeper	3.4.13	3.4.12	3.4.12	3.4.12

5.19.0 release notes

The following release notes include information for Amazon EMR release 5.19.0. Changes are relative to 5.18.0.

Initial release date: November 7, 2018

Last updated date: November 19, 2018

Upgrades

- Hadoop 2.8.5
- Flink 1.6.1
- JupyterHub 0.9.4
- MXNet 1.3.0
- Presto 0.212
- TensorFlow 1.11.0
- Zookeeper 3.4.13
- AWS SDK for Java 1.11.433

New features

- (Nov. 19, 2018) EMR Notebooks is a managed environment based on Jupyter Notebook. It supports Spark magic kernels for PySpark, Spark SQL, Spark R, and Scala. EMR Notebooks can be used with clusters created using Amazon EMR release 5.18.0 and later. For more information, see Using EMR Notebooks in the *Amazon EMR Management Guide*.

- The EMRFS S3-optimized committer is available when writing Parquet files using Spark and EMRFS. This committer improves write performance. For more information, see Use the EMRFS S3-optimized committer.

Changes, enhancements, and resolved issues

- YARN
  - Modified the logic that limits the application master process to running on core nodes. This functionality now uses the YARN node labels feature and properties in the `yarn-site` and `capacity-scheduler` configuration classifications. For information, see https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-instances-guidelines.html#emr-plan-spot-YARN.
- Default Amazon Linux AMI for Amazon EMR
  - `ruby18`, `php56`, and `gcc48` are no longer installed by default. These can be installed if desired using `yum`.
  - The aws-sdk ruby gem is no longer installed by default. It can be installed using `gem install aws-sdk`, if desired. Specific components can also be installed. For example, `gem install aws-sdk-s3`.

Known issues

- **EMR Notebooks**—In some circumstances, with multiple notebook editors open, the notebook editor may appear unable to connect to the cluster. If this happens, clear browser cookies and then reopen notebook editors.
- **CloudWatch ContainerPending Metric and Automatic Scaling**—(Fixed in 5.20.0)Amazon EMR may emit a negative value for `ContainerPending`. If `ContainerPending` is used in an automatic scaling rule, automatic scaling does not behave as expected. Avoid using `ContainerPending` with automatic scaling.
- In Amazon EMR version 5.19.0, 5.20.0, and 5.21.0, YARN node labels are stored in an HDFS directory. In some situations, this leads to core node startup delays and then causes cluster time-out and launch failure. Beginning with Amazon EMR 5.22.0, this issue is resolved. YARN node labels are stored on the local disk of each cluster node, avoiding dependencies on HDFS.

5.19.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.2.0	Amazon SageMaker Spark SDK
emr-ddb	4.7.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.5.1	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.10.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.1.0	EMR S3Select Connector
emrfs	2.28.0	Amazon S3 connector for Hadoop ecosystem applicati ons.

Component	Version	Description
flink-client	1.6.1	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.5-amzn-0	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-0	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.5-amzn-0	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.5-amzn-0	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.5-amzn-0	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.5-amzn-0	Cryptographic key management server based on Hadoop's KeyProvider API.

Component	Version	Description
hadoop-mapred	2.8.5-amzn-0	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-0	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.5-amzn-0	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.5-amzn-0	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.7	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.7	Service for serving one or more HBase regions.
hbase-client	1.4.7	HBase command-line client.
hbase-rest-server	1.4.7	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.7	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.3-amzn-2	The 'hcat' command line client for manipulating hcatalog-server.

Component	Version	Description
hcatalog-server	2.3.3-amzn-2	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.3-amzn-2	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.3-amzn-2	Hive command line client.
hive-hbase	2.3.3-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-2	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.3-amzn-2	Service for accepting Hive queries as web requests.
hue-server	4.2.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.9.4	Multi-user server for Jupyter notebooks
livy-server	0.5.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.

Component	Version	Description
mxnet	1.3.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.0.0	Oozie command-line client.
oozie-server	5.0.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.212	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.212	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.

Component	Version	Description
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.2	Spark command-line clients.
spark-history-server	2.3.2	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.3.2	In-memory execution engine for YARN.
spark-yarn-slave	2.3.2	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	1.11.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.0	Web-based notebook that enables interactive data analytics.

Component	Version	Description
zookeeper-server	3.4.13	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.13	ZooKeeper command line client.

5.19.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.19.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.

Classifications	Description
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.

Classifications	Description
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.

Classifications	Description
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.

Classifications	Description
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.

Classifications	Description
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-memory	Change values in Presto's memory.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
presto-connector-tpcds	Change values in Presto's tpcds.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.

Classifications	Description
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.18.1

5.18.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases

- [Application versions in Amazon EMR 6.x releases]
- [Application versions in Amazon EMR 5.x releases]
- [Application versions in Amazon EMR 4.x releases]

Application version information

	emr-5.18.1	emr-5.18.0	emr-5.17.2	emr-5.17.1
AWS SDK for Java	1.11.393	1.11.393	1.11.336	1.11.336
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.6.0	1.6.0	1.5.2	1.5.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.7	1.4.7	1.4.6	1.4.6
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-

	emr-5.18.1	emr-5.18.0	emr-5.17.2	emr-5.17.1
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.2.0	1.2.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.210	0.210	0.206	0.206
Spark	2.3.2	2.3.2	2.3.1	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.9.0	1.9.0	1.9.0	1.9.0
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

5.18.1 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

Important

In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

5.18.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.1.3	Amazon SageMaker Spark SDK
emr-ddb	4.6.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.5.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.10.0	Distributed copy application optimized for Amazon S3.

Component	Version	Description
emr-s3-select	1.1.0	EMR S3Select Connector
emrfs	2.27.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.6.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.4-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.4-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.4-amzn-1	HTTP endpoint for HDFS operations.

Component	Version	Description
hadoop-kms-server	2.8.4-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.4-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.4-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.7	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.7	Service for serving one or more HBase regions.
hbase-client	1.4.7	HBase command-line client.
hbase-rest-server	1.4.7	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.7	Service providing a Thrift endpoint to HBase.

Component	Version	Description
hcatalog-client	2.3.3-amzn-2	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.3-amzn-2	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.3-amzn-2	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.3-amzn-2	Hive command line client.
hive-hbase	2.3.3-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-2	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.3-amzn-2	Service for accepting Hive queries as web requests.
hue-server	4.2.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.8.1	Multi-user server for Jupyter notebooks
livy-server	0.5.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server

Component	Version	Description
mahout-client	0.13.0	Library for machine learning.
mxnet	1.2.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.0.0	Oozie command-line client.
oozie-server	5.0.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.210	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.210	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.

Component	Version	Description
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.2	Spark command-line clients.
spark-history-server	2.3.2	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.3.2	In-memory execution engine for YARN.
spark-yarn-slave	2.3.2	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	1.9.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.0	Web-based notebook that enables interactive data analytics.

Component	Version	Description
zookeeper-server	3.4.12	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.12	ZooKeeper command line client.

5.18.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.18.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.

Classifications	Description
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.

Classifications	Description
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.

Classifications	Description
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.

Classifications	Description
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.

Classifications	Description
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.

Classifications	Description
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.18.0

5.18.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.18.0	emr-5.17.2	emr-5.17.1	emr-5.17.0
AWS SDK for Java	1.11.393	1.11.336	1.11.336	1.11.336
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.6.0	1.5.2	1.5.2	1.5.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.7	1.4.6	1.4.6	1.4.6
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.2.0	1.2.0	1.2.0	1.2.0

	emr-5.18.0	emr-5.17.2	emr-5.17.1	emr-5.17.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.210	0.206	0.206	0.206
Spark	2.3.2	2.3.1	2.3.1	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.9.0	1.9.0	1.9.0	1.9.0
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

5.18.0 release notes

The following release notes include information for Amazon EMR release 5.18.0. Changes are relative to 5.17.0.

Initial release date: October 24, 2018

Upgrades

- Flink 1.6.0
- HBase 1.4.7
- Presto 0.210
- Spark 2.3.2

- Zeppelin 0.8.0

New features

- Beginning with Amazon EMR 5.18.0, you can use the Amazon EMR artifact repository to build your job code against the exact versions of libraries and dependencies that are available with specific Amazon EMR releases. For more information, see Checking dependencies using the Amazon EMR artifact repository.

Changes, enhancements, and resolved issues

- Hive
  - Added support for S3 Select. For more information, see Using S3 Select with Hive to improve performance.
- Presto
  - Added support for S3 Select Pushdown. For more information, see Using S3 Select Pushdown with Presto to improve performance.
- Spark
  - The default log4j configuration for Spark has been changed to roll container logs hourly for Spark streaming jobs. This helps prevent the deletion of logs for long-running Spark streaming jobs.

5.18.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.1.3	Amazon SageMaker Spark SDK
emr-ddb	4.6.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.5.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.10.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.1.0	EMR S3Select Connector
emrfs	2.27.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.6.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.

Component	Version	Description
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.4-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.4-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.4-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.4-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.4-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.4-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.

Component	Version	Description
hadoop-yarn-timeline-server	2.8.4-amzn-1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.7	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.7	Service for serving one or more HBase regions.
hbase-client	1.4.7	HBase command-line client.
hbase-rest-server	1.4.7	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.7	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.3-amzn-2	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.3-amzn-2	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.3-amzn-2	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.3-amzn-2	Hive command line client.
hive-hbase	2.3.3-amzn-2	Hive-hbase client.

Component	Version	Description
hive-metastore-server	2.3.3-amzn-2	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.3-amzn-2	Service for accepting Hive queries as web requests.
hue-server	4.2.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.8.1	Multi-user server for Jupyter notebooks
livy-server	0.5.0-incubating	REST interface for interacting with Apache Spark
nginx	1.12.1	nginx [engine x] is an HTTP and reverse proxy server
mahout-client	0.13.0	Library for machine learning.
mxnet	1.2.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.0.0	Oozie command-line client.
oozie-server	5.0.0	Service for accepting Oozie workflow requests.

Component	Version	Description
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.210	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.210	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.2	Spark command-line clients.
spark-history-server	2.3.2	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.3.2	In-memory execution engine for YARN.
spark-yarn-slave	2.3.2	Apache Spark libraries needed by YARN slaves.

Component	Version	Description
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	1.9.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.8.0	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.12	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.12	ZooKeeper command line client.

5.18.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.18.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.

Classifications	Description
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.

Classifications	Description
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.

Classifications	Description
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.

Classifications	Description
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.

Classifications	Description
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.17.2

5.17.2 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.17.2	emr-5.17.1	emr-5.17.0	emr-5.16.1
AWS SDK for Java	1.11.336	1.11.336	1.11.336	1.11.336
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.5.2	1.5.2	1.5.2	1.5.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.6	1.4.6	1.4.6	1.4.4

	emr-5.17.2	emr-5.17.1	emr-5.17.0	emr-5.16.1
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.2.0	1.2.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.206	0.206	0.206	0.203
Spark	2.3.1	2.3.1	2.3.1	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.9.0	1.9.0	1.9.0	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.17.2	emr-5.17.1	emr-5.17.0	emr-5.16.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

5.17.2 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

Important

In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

5.17.2 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.1.3	Amazon SageMaker Spark SDK

Component	Version	Description
emr-ddb	4.6.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.5.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.10.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.0.0	EMR S3Select Connector
emrfs	2.26.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.5.2	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.

Component	Version	Description
hadoop-client	2.8.4-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.4-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.4-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.4-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.4-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.4-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Service for retrieving current and historical information for YARN applications.

Component	Version	Description
hbase-hmaster	1.4.6	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.6	Service for serving one or more HBase regions.
hbase-client	1.4.6	HBase command-line client.
hbase-rest-server	1.4.6	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.6	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.3-amzn-1	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.3-amzn-1	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.3-amzn-1	Hive command line client.
hive-hbase	2.3.3-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.

Component	Version	Description
hive-server2	2.3.3-amzn-1	Service for accepting Hive queries as web requests.
hue-server	4.2.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.8.1	Multi-user server for Jupyter notebooks
livy-server	0.5.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.
mxnet	1.2.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.0.0	Oozie command-line client.
oozie-server	5.0.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	The phoenix libraries for server and client

Component	Version	Description
phoenix-query-server	4.14.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.206	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.206	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.1	Spark command-line clients.
spark-history-server	2.3.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.3.1	In-memory execution engine for YARN.
spark-yarn-slave	2.3.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.

Component	Version	Description
tensorflow	1.9.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.3	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.12	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.12	ZooKeeper command line client.

5.17.2 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.17.2 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.

Classifications	Description
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.

Classifications	Description
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.

Classifications	Description
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.

Classifications	Description
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.

Classifications	Description
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.

Classifications	Description
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.17.1

5.17.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.17.1	**emr-5.17.0**	**emr-5.16.1**	**emr-5.16.0**
AWS SDK for Java	1.11.336	1.11.336	1.11.336	1.11.336
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.5.2	1.5.2	1.5.0	1.5.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.6	1.4.6	1.4.4	1.4.4
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-

	emr-5.17.1	emr-5.17.0	emr-5.16.1	emr-5.16.0
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.2.0	1.2.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.206	0.206	0.203	0.203
Spark	2.3.1	2.3.1	2.3.1	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.9.0	1.9.0	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

5.17.1 release notes

The following release notes include information for Amazon EMR release 5.17.1. Changes are relative to 5.17.0.

Initial release date: July 18, 2019

Changes, enhancements, and resolved issues

- Updated the default Amazon Linux AMI for Amazon EMR to include important Linux kernel security updates, including the TCP SACK Denial of Service Issue (AWS-2019-005).

5.17.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.1.3	Amazon SageMaker Spark SDK
emr-ddb	4.6.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.5.0	Extra convenience libraries for the Hadoop ecosystem.

Component	Version	Description
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.10.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.0.0	EMR S3Select Connector
emrfs	2.26.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.5.2	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.4-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFS node-level service for storing blocks.

Component	Version	Description
hadoop-hdfs-library	2.8.4-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.4-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.4-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.4-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.4-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.6	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.6	Service for serving one or more HBase regions.

Component	Version	Description
hbase-client	1.4.6	HBase command-line client.
hbase-rest-server	1.4.6	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.6	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.3-amzn-1	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.3-amzn-1	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.3-amzn-1	Hive command line client.
hive-hbase	2.3.3-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.3-amzn-1	Service for accepting Hive queries as web requests.
hue-server	4.2.0	Web application for analyzing data using Hadoop ecosystem applications

Component	Version	Description
jupyterhub	0.8.1	Multi-user server for Jupyter notebooks
livy-server	0.5.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.
mxnet	1.2.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.0.0	Oozie command-line client.
oozie-server	5.0.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API

Component	Version	Description
presto-coordinator	0.206	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.206	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.1	Spark command-line clients.
spark-history-server	2.3.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.3.1	In-memory execution engine for YARN.
spark-yarn-slave	2.3.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	1.9.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.

Component	Version	Description
zeppelin-server	0.7.3	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.12	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.12	ZooKeeper command line client.

5.17.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.17.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.

Classifications	Description
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.

Classifications	Description
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.

Classifications	Description
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.

Classifications	Description
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.

Classifications	Description
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.

Classifications	Description
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.17.0

5.17.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, TensorFlow, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases

- [Application versions in Amazon EMR 4.x releases]

Application version information

	emr-5.17.0	emr-5.16.1	emr-5.16.0	emr-5.15.1
AWS SDK for Java	1.11.336	1.11.336	1.11.336	1.11.333
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.5.2	1.5.0	1.5.0	1.4.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.6	1.4.4	1.4.4	1.4.4
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.3
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1

	emr-5.17.0	emr-5.16.1	emr-5.16.0	emr-5.15.1
Livy	0.5.0	0.5.0	0.5.0	0.4.0
MXNet	1.2.0	1.2.0	1.2.0	1.1.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.206	0.203	0.203	0.194
Spark	2.3.1	2.3.1	2.3.1	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.9.0	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

5.17.0 release notes

The following release notes include information for Amazon EMR release 5.17.0. Changes are relative to 5.16.0.

Initial release date: August 30, 2018

Upgrades

- Flink 1.5.2

- HBase 1.4.6
- Presto 0.206

New features

- Added support for Tensorflow. For more information, see TensorFlow.

Changes, enhancements, and resolved issues

- JupyterHub
  - Added support for notebook persistence in Amazon S3. For more information, see Configuring persistence for notebooks in Amazon S3.
- Spark
  - Added support for S3 Select. For more information, see Use S3 Select with Spark to improve query performance.
- Resolved the issues with the Cloudwatch metrics and the automatic scaling feature in Amazon EMR version 5.14.0, 5.15.0, or 5.16.0.

Known issues

- When you create a kerberized cluster with Livy installed, Livy fails with an error that simple authentication is not enabled. Rebooting the Livy server resolves the issue. As a workaround, add a step during cluster creation that runs `sudo restart livy-server` on the primary node.
- If you use a custom Amazon Linux AMI based on an Amazon Linux AMI with a creation date of 2018-08-11, the Oozie server fails to start. If you use Oozie, create a custom AMI based on an Amazon Linux AMI ID with a different creation date. You can use the following AWS CLI command to return a list of Image IDs for all HVM Amazon Linux AMIs with a 2018.03 version, along with the release date, so that you can choose an appropriate Amazon Linux AMI as your base. Replace MyRegion with your Region identifier, such as us-west-2.

```
aws ec2 --region MyRegion describe-images --owner amazon --query 'Images[?
Name!=`null`]|[?starts_with(Name, `amzn-ami-hvm-2018.03`) == `true`].
[CreationDate,ImageId,Name]' --output text | sort -rk1
```

5.17.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.1.3	Amazon SageMaker Spark SDK
emr-ddb	4.6.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.5.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.10.0	Distributed copy application optimized for Amazon S3.
emr-s3-select	1.0.0	EMR S3Select Connector
emrfs	2.26.0	Amazon S3 connector for Hadoop ecosystem applicati ons.

Component	Version	Description
flink-client	1.5.2	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.4-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.4-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.4-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.4-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.

Component	Version	Description
hadoop-mapred	2.8.4-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.4-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.6	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.6	Service for serving one or more HBase regions.
hbase-client	1.4.6	HBase command-line client.
hbase-rest-server	1.4.6	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.6	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.3-amzn-1	The 'hcat' command line client for manipulating hcatalog-server.

Component	Version	Description
hcatalog-server	2.3.3-amzn-1	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.3-amzn-1	Hive command line client.
hive-hbase	2.3.3-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.3-amzn-1	Service for accepting Hive queries as web requests.
hue-server	4.2.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.8.1	Multi-user server for Jupyter notebooks
livy-server	0.5.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.
mxnet	1.2.0	A flexible, scalable, and efficient library for deep learning.

Component	Version	Description
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.0.0	Oozie command-line client.
oozie-server	5.0.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.206	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.206	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.1	Spark command-line clients.

Component	Version	Description
spark-history-server	2.3.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.3.1	In-memory execution engine for YARN.
spark-yarn-slave	2.3.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tensorflow	1.9.0	TensorFlow open source software library for high performance numerical computation.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.3	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.12	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.12	ZooKeeper command line client.

5.17.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.17.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.

Classifications	Description
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.

Classifications	Description
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-s3-conf	Configure Jupyter Notebook S3 persistence.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.

Classifications	Description
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.

Classifications	Description
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.

Classifications	Description
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.16.1

5.16.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.16.1	emr-5.16.0	emr-5.15.1	emr-5.15.0
AWS SDK for Java	1.11.336	1.11.336	1.11.333	1.11.333
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.5.0	1.5.0	1.4.2	1.4.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.4	1.4.4	1.4.4	1.4.4

	emr-5.16.1	emr-5.16.0	emr-5.15.1	emr-5.15.0
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.3	2.8.3
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.4.0	0.4.0
MXNet	1.2.0	1.2.0	1.1.0	1.1.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.203	0.203	0.194	0.194
Spark	2.3.1	2.3.1	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.16.1	emr-5.16.0	emr-5.15.1	emr-5.15.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

5.16.1 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

> **Important**
>
> In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

5.16.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.1.0	Amazon SageMaker Spark SDK

Component	Version	Description
emr-ddb	4.6.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.10.0	Distributed copy application optimized for Amazon S3.
emrfs	2.25.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.5.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.4-amzn-0	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.

Component	Version	Description
hadoop-hdfs-datanode	2.8.4-amzn-0	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.4-amzn-0	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.4-amzn-0	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.4-amzn-0	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.4-amzn-0	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.4-amzn-0	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.4-amzn-0	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.4-amzn-0	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.4-amzn-0	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.4	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.

Component	Version	Description
hbase-region-server	1.4.4	Service for serving one or more HBase regions.
hbase-client	1.4.4	HBase command-line client.
hbase-rest-server	1.4.4	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.4	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.3-amzn-1	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.3-amzn-1	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.3-amzn-1	Hive command line client.
hive-hbase	2.3.3-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.3-amzn-1	Service for accepting Hive queries as web requests.

Component	Version	Description
hue-server	4.2.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.8.1	Multi-user server for Jupyter notebooks
livy-server	0.5.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.
mxnet	1.2.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.0.0	Oozie command-line client.
oozie-server	5.0.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API

Component	Version	Description
presto-coordinator	0.203	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.203	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.1	Spark command-line clients.
spark-history-server	2.3.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.3.1	In-memory execution engine for YARN.
spark-yarn-slave	2.3.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.3	Web-based notebook that enables interactive data analytics.

Component	Version	Description
zookeeper-server	3.4.12	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.12	ZooKeeper command line client.

5.16.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.16.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.

Classifications	Description
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.

Classifications	Description
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.

Classifications	Description
hadoop-kms-env	Change values in the Hadoop KMS environment.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhub_config.py file.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.

Classifications	Description
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.

Classifications	Description
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.

Classifications	Description
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.16.0

5.16.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.16.0	emr-5.15.1	emr-5.15.0	emr-5.14.2
AWS SDK for Java	1.11.336	1.11.333	1.11.333	1.11.297
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.5.0	1.4.2	1.4.2	1.4.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.4	1.4.4	1.4.4	1.4.2
HCatalog	2.3.3	2.3.3	2.3.3	2.3.2
Hadoop	2.8.4	2.8.3	2.8.3	2.8.3
Hive	2.3.3	2.3.3	2.3.3	2.3.2
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.4.0	0.4.0	0.4.0
MXNet	1.2.0	1.1.0	1.1.0	1.1.0

	emr-5.16.0	emr-5.15.1	emr-5.15.0	emr-5.14.2
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	4.3.0
Phoenix	4.14.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.203	0.194	0.194	0.194
Spark	2.3.1	2.3.0	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.10

5.16.0 release notes

The following release notes include information for Amazon EMR release 5.16.0. Changes are relative to 5.15.0.

Initial release date: July 19, 2018

Upgrades

- Hadoop 2.8.4
- Flink 1.5.0
- Livy 0.5.0
- MXNet 1.2.0

- Phoenix 4.14.0
- Presto 0.203
- Spark 2.3.1
- AWS SDK for Java 1.11.336
- CUDA 9.2
- Redshift JDBC Driver 1.2.15.1025

Changes, enhancements, and resolved issues

- HBase
  - Backported HBASE-20723
- Presto
  - Configuration changes to support LDAP authentication. For more information, see Using LDAP authentication for Presto on Amazon EMR.
- Spark
  - Apache Spark version 2.3.1, available beginning with Amazon EMR release 5.16.0, addresses CVE-2018-8024 and CVE-2018-1334. We recommend that you migrate earlier versions of Spark to Spark version 2.3.1 or later.

Known issues

- This release version does not support the c1.medium or m1.small instance types. Clusters using either of these instance types fail to start. As a workaround, specify a different instance type or use a different release version.
- When you create a kerberized cluster with Livy installed, Livy fails with an error that simple authentication is not enabled. Rebooting the Livy server resolves the issue. As a workaround, add a step during cluster creation that runs `sudo restart livy-server` on the primary node.
- After the primary node reboots or the instance controller restarts, the CloudWatch metrics will not be collected and the automatic scaling feature will not be available in Amazon EMR version 5.14.0, 5.15.0, or 5.16.0. This issue is fixed in Amazon EMR 5.17.0.

5.16.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.1.0	Amazon SageMaker Spark SDK
emr-ddb	4.6.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.10.0	Distributed copy application optimized for Amazon S3.
emrfs	2.25.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.5.0	Apache Flink command line client scripts and applications.

Component	Version	Description
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.4-amzn-0	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.4-amzn-0	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.4-amzn-0	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.4-amzn-0	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.4-amzn-0	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.4-amzn-0	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.4-amzn-0	MapReduce execution engine libraries for running a MapReduce application.

Component	Version	Description
hadoop-yarn-nodemanager	2.8.4-amzn-0	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.4-amzn-0	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.4-amzn-0	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.4	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.4	Service for serving one or more HBase regions.
hbase-client	1.4.4	HBase command-line client.
hbase-rest-server	1.4.4	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.4	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.3-amzn-1	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.3-amzn-1	Service providing HCatalog, a table and storage management layer for distributed applications.

Component	Version	Description
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.3-amzn-1	Hive command line client.
hive-hbase	2.3.3-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.3-amzn-1	Service for accepting Hive queries as web requests.
hue-server	4.2.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.8.1	Multi-user server for Jupyter notebooks
livy-server	0.5.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.
mxnet	1.2.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.2.88	Nvidia drivers and Cuda toolkit
oozie-client	5.0.0	Oozie command-line client.

Component	Version	Description
oozie-server	5.0.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.14.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.203	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.203	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.1	Spark command-line clients.
spark-history-server	2.3.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.3.1	In-memory execution engine for YARN.

Component	Version	Description
spark-yarn-slave	2.3.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.3	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.12	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.12	ZooKeeper command line client.

5.16.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.16.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.

Classifications	Description
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.

Classifications	Description
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.

Classifications	Description
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environment.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhub_config.py file.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.

Classifications	Description
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-password-authenticator	Change values in Presto's password-authentic ator.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.

Classifications	Description
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.

Classifications	Description
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.15.1

5.15.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.15.1	emr-5.15.0	emr-5.14.2	emr-5.14.1
AWS SDK for Java	1.11.333	1.11.333	1.11.297	1.11.297
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.2	1.4.2	1.4.2	1.4.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.4	1.4.4	1.4.2	1.4.2
HCatalog	2.3.3	2.3.3	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.3	2.3.3	2.3.2	2.3.2
Hudi	-	-	-	-

	emr-5.15.1	emr-5.15.0	emr-5.14.2	emr-5.14.1
Hue	4.2.0	4.2.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.1.0	1.1.0	1.1.0	1.1.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.194	0.194
Spark	2.3.0	2.3.0	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.10	3.4.10

5.15.1 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

> **Important**
>
> In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

5.15.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.

Component	Version	Description
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.10.0	Distributed copy application optimized for Amazon S3.
emrfs	2.24.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.4.2	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.3-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.3-amzn-1	HDFS command-line client and library

Component	Version	Description
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.3-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.3-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.3-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.3-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.4	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.4	Service for serving one or more HBase regions.
hbase-client	1.4.4	HBase command-line client.

Component	Version	Description
hbase-rest-server	1.4.4	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.4	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.3-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.3-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.3-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.3-amzn-0	Hive command line client.
hive-hbase	2.3.3-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.3-amzn-0	Service for accepting Hive queries as web requests.
hue-server	4.2.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.8.1	Multi-user server for Jupyter notebooks

Component	Version	Description
livy-server	0.4.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.
mxnet	1.1.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.1.85	Nvidia drivers and Cuda toolkit
oozie-client	5.0.0	Oozie command-line client.
oozie-server	5.0.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.13.0-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.13.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.194	Service for accepting queries and managing query execution among presto-wo rkers.

Component	Version	Description
presto-worker	0.194	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.0	Spark command-line clients.
spark-history-server	2.3.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.3.0	In-memory execution engine for YARN.
spark-yarn-slave	2.3.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.3	Web-based notebook that enables interactive data analytics.

Component	Version	Description
zookeeper-server	3.4.12	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.12	ZooKeeper command line client.

5.15.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.15.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.

Classifications	Description
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.

Classifications	Description
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.

Classifications	Description
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.

Classifications	Description
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.

Classifications	Description
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.

Classifications	Description
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.15.0

5.15.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.15.0	emr-5.14.2	emr-5.14.1	emr-5.14.0
AWS SDK for Java	1.11.333	1.11.297	1.11.297	1.11.297

	emr-5.15.0	emr-5.14.2	emr-5.14.1	emr-5.14.0
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.2	1.4.2	1.4.2	1.4.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.4	1.4.2	1.4.2	1.4.2
HCatalog	2.3.3	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.3	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.2.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.1.0	1.1.0	1.1.0	1.1.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	4.3.0	4.3.0	4.3.0

	emr-5.15.0	emr-5.14.2	emr-5.14.1	emr-5.14.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.194	0.194
Spark	2.3.0	2.3.0	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.10	3.4.10	3.4.10

5.15.0 release notes

The following release notes include information for Amazon EMR release 5.15.0. Changes are relative to 5.14.0.

Initial release date: June 21, 2018

Upgrades

- Upgraded HBase to 1.4.4
- Upgraded Hive to 2.3.3
- Upgraded Hue to 4.2.0
- Upgraded Oozie to 5.0.0
- Upgraded Zookeeper to 3.4.12
- Upgraded AWS SDK to 1.11.333

Changes, enhancements, and resolved issues

- Hive
  - Backported HIVE-18069
- Hue
  - Updated Hue to correctly authenticate with Livy when Kerberos is enabled. Livy is now supported when using Kerberos with Amazon EMR.
- JupyterHub
  - Updated JupyterHub so that Amazon EMR installs LDAP client libraries by default.
  - Fixed an error in the script that generates self-signed certificates.

Known issues

- This release version does not support the c1.medium or m1.small instance types. Clusters using either of these instance types fail to start. As a workaround, specify a different instance type or use a different release version.
- After the primary node reboots or the instance controller restarts, the CloudWatch metrics will not be collected and the automatic scaling feature will not be available in Amazon EMR version 5.14.0, 5.15.0, or 5.16.0. This issue is fixed in Amazon EMR 5.17.0.

5.15.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.10.0	Distributed copy application optimized for Amazon S3.
emrfs	2.24.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.4.2	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.

Component	Version	Description
hadoop-client	2.8.3-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.3-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.3-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.3-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.3-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.3-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Service for retrieving current and historical information for YARN applications.

Component	Version	Description
hbase-hmaster	1.4.4	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.4	Service for serving one or more HBase regions.
hbase-client	1.4.4	HBase command-line client.
hbase-rest-server	1.4.4	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.4	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.3-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.3-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.3-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.3-amzn-0	Hive command line client.
hive-hbase	2.3.3-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.

Component	Version	Description
hive-server2	2.3.3-amzn-0	Service for accepting Hive queries as web requests.
hue-server	4.2.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.8.1	Multi-user server for Jupyter notebooks
livy-server	0.4.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.
mxnet	1.1.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.1.85	Nvidia drivers and Cuda toolkit
oozie-client	5.0.0	Oozie command-line client.
oozie-server	5.0.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.13.0-HBase-1.4	The phoenix libraries for server and client

Component	Version	Description
phoenix-query-server	4.13.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.194	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.194	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.0	Spark command-line clients.
spark-history-server	2.3.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.3.0	In-memory execution engine for YARN.
spark-yarn-slave	2.3.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.

Component	Version	Description
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.3	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.12	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.12	ZooKeeper command line client.

5.15.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.15.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.

Classifications	Description
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.

Classifications	Description
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.

Classifications	Description
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.

Classifications	Description
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.

Classifications	Description
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.

Classifications	Description
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.14.2

5.14.2 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.14.2	emr-5.14.1	emr-5.14.0	emr-5.13.1
AWS SDK for Java	1.11.297	1.11.297	1.11.297	1.11.297
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.2	1.4.2	1.4.2	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.2	1.4.2	1.4.2	1.4.2
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.1.0	1.1.0	1.1.0	1.0.0

	emr-5.14.2	emr-5.14.1	emr-5.14.0	emr-5.13.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.194	0.194
Spark	2.3.0	2.3.0	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.14.2 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

Important

In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

5.14.2 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.10.0	Distributed copy application optimized for Amazon S3.
emrfs	2.23.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.4.2	Apache Flink command line client scripts and applications.

Component	Version	Description
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.3-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.3-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.3-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.3-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.3-amzn-1	MapReduce execution engine libraries for running a MapReduce application.

Component	Version	Description
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.3-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.2	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.2	Service for serving one or more HBase regions.
hbase-client	1.4.2	HBase command-line client.
hbase-rest-server	1.4.2	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.2	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.2-amzn-2	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.2-amzn-2	Service providing HCatalog, a table and storage management layer for distributed applications.

Component	Version	Description
hcatalog-webhcat-server	2.3.2-amzn-2	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.2-amzn-2	Hive command line client.
hive-hbase	2.3.2-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-2	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.2-amzn-2	Service for accepting Hive queries as web requests.
hue-server	4.1.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.8.1	Multi-user server for Jupyter notebooks
livy-server	0.4.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.
mxnet	1.1.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.1.85	Nvidia drivers and Cuda toolkit
oozie-client	4.3.0	Oozie command-line client.

Component	Version	Description
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.13.0-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.13.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.194	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.194	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.0	Spark command-line clients.
spark-history-server	2.3.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.3.0	In-memory execution engine for YARN.

Component	Version	Description
spark-yarn-slave	2.3.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.3	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.14.2 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.14.2 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.

Classifications	Description
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.

Classifications	Description
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.

Classifications	Description
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.

Classifications	Description
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.

Classifications	Description
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file

Classifications	Description
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.14.1

5.14.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- [Application versions in Amazon EMR 7.x releases]
- [Application versions in Amazon EMR 6.x releases]
- [Application versions in Amazon EMR 5.x releases]
- [Application versions in Amazon EMR 4.x releases]

Application version information

	emr-5.14.1	emr-5.14.0	emr-5.13.1	emr-5.13.0
AWS SDK for Java	1.11.297	1.11.297	1.11.297	1.11.297
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.2	1.4.2	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.2	1.4.2	1.4.2	1.4.2
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-

	emr-5.14.1	emr-5.14.0	emr-5.13.1	emr-5.13.0
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.1.0	1.1.0	1.0.0	1.0.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.194	0.194
Spark	2.3.0	2.3.0	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.14.1 release notes

The following release notes include information for Amazon EMR release 5.14.1. Changes are relative to 5.14.0.

Initial release date: October 17, 2018

Updated the default AMI for Amazon EMR to address potential security vulnerabilities.

5.14.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.10.0	Distributed copy application optimized for Amazon S3.

Component	Version	Description
emrfs	2.23.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.4.2	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.3-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.3-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.3-amzn-1	HTTP endpoint for HDFS operations.

Component	Version	Description
hadoop-kms-server	2.8.3-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.3-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.3-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.2	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.2	Service for serving one or more HBase regions.
hbase-client	1.4.2	HBase command-line client.
hbase-rest-server	1.4.2	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.2	Service providing a Thrift endpoint to HBase.

Component	Version	Description
hcatalog-client	2.3.2-amzn-2	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.2-amzn-2	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.2-amzn-2	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.2-amzn-2	Hive command line client.
hive-hbase	2.3.2-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-2	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.2-amzn-2	Service for accepting Hive queries as web requests.
hue-server	4.1.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.8.1	Multi-user server for Jupyter notebooks
livy-server	0.4.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.

Component	Version	Description
mxnet	1.1.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.1.85	Nvidia drivers and Cuda toolkit
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.13.0-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.13.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.194	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.194	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.

Component	Version	Description
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.0	Spark command-line clients.
spark-history-server	2.3.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.3.0	In-memory execution engine for YARN.
spark-yarn-slave	2.3.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.3	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.14.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see [Configure applications](#).

emr-5.14.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.

Classifications	Description
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.

Classifications	Description
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.

Classifications	Description
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.

Classifications	Description
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.

Classifications	Description
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.14.0

5.14.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, JupyterHub, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.14.0	emr-5.13.1	emr-5.13.0	emr-5.12.3
AWS SDK for Java	1.11.297	1.11.297	1.11.297	1.11.267
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.2	1.4.0	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.2	1.4.2	1.4.2	1.4.0

	emr-5.14.0	emr-5.13.1	emr-5.13.0	emr-5.12.3
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.1.0	1.0.0	1.0.0	1.0.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.194	0.188
Spark	2.3.0	2.3.0	2.3.0	2.2.1
Sqoop	1.4.7	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.14.0	emr-5.13.1	emr-5.13.0	emr-5.12.3
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.14.0 release notes

The following release notes include information for Amazon EMR release 5.14.0. Changes are relative to 5.13.0.

Initial release date: June 4, 2018

Upgrades

- Upgraded Apache Flink to 1.4.2
- Upgraded Apache MXnet to 1.1.0
- Upgraded Apache Sqoop to 1.4.7

New features

- Added JupyterHub support. For more information, see JupyterHub.

Changes, enhancements, and resolved issues

- EMRFS
  - The userAgent string in requests to Amazon S3 has been updated to contain the user and group information of the invoking principal. This can be used with AWS CloudTrail logs for more comprehensive request tracking.
- HBase
  - Included HBASE-20447, which addresses an issue that could cause cache issues, especially with split Regions.
- MXnet

  - Added OpenCV libraries.
- Spark
  - When Spark writes Parquet files to an Amazon S3 location using EMRFS, the FileOutputCommitter algorithm has been updated to use version 2 instead of version 1. This reduces the number of renames, which improves application performance. This change does not affect:
    - Applications other than Spark.
    - Applications that write to other file systems, such as HDFS (which still use version 1 of FileOutputCommitter).
    - Applications that use other output formats, such as text or csv, that already use EMRFS direct write.

Known issues

- JupyterHub
  - Using configuration classifications to set up JupyterHub and individual Jupyter notebooks when you create a cluster is not supported. Edit the jupyterhub_config.py file and jupyter_notebook_config.py files for each user manually. For more information, see Configuring JupyterHub.
  - JupyterHub fails to start on clusters within a private subnet, failing with the message `Error: ENOENT: no such file or directory, open '/etc/jupyter/conf/server.crt'` . This is caused by an error in the script that generates self-signed certificates. Use the following workaround to generate self-signed certificates. All commands are executed while connected to the primary node.
    1. Copy the certificate generation script from the container to the primary node:

       ```
       sudo docker cp jupyterhub:/tmp/gen_self_signed_cert.sh ./
       ```

    2. Use a text editor to change line 23 to change public hostname to local hostname as shown below:

       ```
       local hostname=$(curl -s $EC2_METADATA_SERVICE_URI/local-hostname)
       ```

    3. Run the script to generate self-signed certificates:

```
sudo bash ./gen_self_signed_cert.sh
```

4. Move the certificate files that the script generates to the `/etc/jupyter/conf/` directory:

```
sudo mv /tmp/server.crt /tmp/server.key /etc/jupyter/conf/
```

You can `tail` the `jupyter.log` file to verify that JupyterHub restarted and is returning a 200 response code. For example:

```
tail -f /var/log/jupyter/jupyter.log
```

This should return a response similar to the following:

```
# [I 2018-06-14 18:56:51.356 JupyterHub app:1581] JupyterHub is now running at
 https://:9443/
# 19:01:51.359 - info: [ConfigProxy] 200 GET /api/routes
```

- After the primary node reboots or the instance controller restarts, the CloudWatch metrics will not be collected and the automatic scaling feature will not be available in Amazon EMR version 5.14.0, 5.15.0, or 5.16.0. This issue is fixed in Amazon EMR 5.17.0.

5.14.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.10.0	Distributed copy application optimized for Amazon S3.
emrfs	2.23.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.4.2	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.

Component	Version	Description
hadoop-client	2.8.3-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.3-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.3-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.3-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.3-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.3-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Service for retrieving current and historical information for YARN applications.

Component	Version	Description
hbase-hmaster	1.4.2	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.2	Service for serving one or more HBase regions.
hbase-client	1.4.2	HBase command-line client.
hbase-rest-server	1.4.2	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.2	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.2-amzn-2	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.2-amzn-2	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.2-amzn-2	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.2-amzn-2	Hive command line client.
hive-hbase	2.3.2-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-2	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.

Component	Version	Description
hive-server2	2.3.2-amzn-2	Service for accepting Hive queries as web requests.
hue-server	4.1.0	Web application for analyzing data using Hadoop ecosystem applications
jupyterhub	0.8.1	Multi-user server for Jupyter notebooks
livy-server	0.4.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.
mxnet	1.1.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.1.85	Nvidia drivers and Cuda toolkit
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.13.0-HBase-1.4	The phoenix libraries for server and client

Component	Version	Description
phoenix-query-server	4.13.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.194	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.194	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.0	Spark command-line clients.
spark-history-server	2.3.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.3.0	In-memory execution engine for YARN.
spark-yarn-slave	2.3.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.7	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.

Component	Version	Description
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.3	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.14.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.14.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
container-log4j	Change values in Hadoop YARN's container-log4j.properties file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.

Classifications	Description
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.

Classifications	Description
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.

Classifications	Description
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
jupyter-notebook-conf	Change values in Jupyter Notebook's jupyter_notebook_config.py file.
jupyter-hub-conf	Change values in JupyterHubs's jupyterhu b_config.py file.
jupyter-sparkmagic-conf	Change values in Sparkmagic's config.json file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.

Classifications	Description
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.

Classifications	Description
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.

Classifications	Description
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.13.1

5.13.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.13.1	emr-5.13.0	emr-5.12.3	emr-5.12.2
AWS SDK for Java	1.11.297	1.11.297	1.11.267	1.11.267
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.2	1.4.2	1.4.0	1.4.0
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.0.0	1.0.0	1.0.0	1.0.0

	emr-5.13.1	emr-5.13.0	emr-5.12.3	emr-5.12.2
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.188	0.188
Spark	2.3.0	2.3.0	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.13.1 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

> **Important**
>
> In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

5.13.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form `CommunityVersion-amzn-EmrVersion`. The `EmrVersion` starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.10.0	Distributed copy application optimized for Amazon S3.
emrfs	2.22.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.4.0	Apache Flink command line client scripts and applications.

Component	Version	Description
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.3-amzn-0	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.3-amzn-0	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.3-amzn-0	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.3-amzn-0	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.3-amzn-0	MapReduce execution engine libraries for running a MapReduce application.

Component	Version	Description
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.3-amzn-0	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.2	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.2	Service for serving one or more HBase regions.
hbase-client	1.4.2	HBase command-line client.
hbase-rest-server	1.4.2	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.2	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.2-amzn-2	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.2-amzn-2	Service providing HCatalog, a table and storage management layer for distributed applications.

Component	Version	Description
hcatalog-webhcat-server	2.3.2-amzn-2	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.2-amzn-2	Hive command line client.
hive-hbase	2.3.2-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-2	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.2-amzn-2	Service for accepting Hive queries as web requests.
hue-server	4.1.0	Web application for analyzing data using Hadoop ecosystem applications
livy-server	0.4.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.
mxnet	1.0.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.1.85	Nvidia drivers and Cuda toolkit
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.

Component	Version	Description
phoenix-library	4.13.0-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.13.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.194	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.194	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.0	Spark command-line clients.
spark-history-server	2.3.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.3.0	In-memory execution engine for YARN.
spark-yarn-slave	2.3.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.

Component	Version	Description
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.3	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.13.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.13.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.

Classifications	Description
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.

Classifications	Description
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.

Classifications	Description
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.

Classifications	Description
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.

Classifications	Description
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.

Classifications	Description
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.13.0

5.13.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.13.0	emr-5.12.3	emr-5.12.2	emr-5.12.1
AWS SDK for Java	1.11.297	1.11.267	1.11.267	1.11.267
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4

	emr-5.13.0	emr-5.12.3	emr-5.12.2	emr-5.12.1
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.2	1.4.0	1.4.0	1.4.0
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.0.0	1.0.0	1.0.0	1.0.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0

	emr-5.13.0	emr-5.12.3	emr-5.12.2	emr-5.12.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.188	0.188	0.188
Spark	2.3.0	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.13.0 release notes

The following release notes include information for the Amazon EMR release 5.13.0. Changes are relative to 5.12.0.

Upgrades

- Upgraded Spark to 2.3.0
- Upgraded HBase to 1.4.2
- Upgraded Presto to 0.194
- Upgraded AWS SDK for Java to 1.11.297

Changes, enhancements, and resolved issues

- Hive
  - Backported HIVE-15436. Enhanced Hive APIs to return only views.

Known issues

- MXNet does not currently have OpenCV libraries.

5.13.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.10.0	Distributed copy application optimized for Amazon S3.

Component	Version	Description
emrfs	2.22.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.4.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.3-amzn-0	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.3-amzn-0	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.3-amzn-0	HTTP endpoint for HDFS operations.

Component	Version	Description
hadoop-kms-server	2.8.3-amzn-0	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.3-amzn-0	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.3-amzn-0	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.2	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.2	Service for serving one or more HBase regions.
hbase-client	1.4.2	HBase command-line client.
hbase-rest-server	1.4.2	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.2	Service providing a Thrift endpoint to HBase.

Component	Version	Description
hcatalog-client	2.3.2-amzn-2	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.2-amzn-2	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.2-amzn-2	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.2-amzn-2	Hive command line client.
hive-hbase	2.3.2-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-2	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.2-amzn-2	Service for accepting Hive queries as web requests.
hue-server	4.1.0	Web application for analyzing data using Hadoop ecosystem applications
livy-server	0.4.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.
mxnet	1.0.0	A flexible, scalable, and efficient library for deep learning.

Component	Version	Description
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.1.85	Nvidia drivers and Cuda toolkit
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.13.0-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.13.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.194	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.194	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.0	Spark command-line clients.
spark-history-server	2.3.0	Web UI for viewing logged events for the lifetime of a completed Spark application.

Component	Version	Description
spark-on-yarn	2.3.0	In-memory execution engine for YARN.
spark-yarn-slave	2.3.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.3	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.13.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.13.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.

Classifications	Description
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.

Classifications	Description
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.

Classifications	Description
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.

Classifications	Description
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.

Classifications	Description
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.12.3

5.12.3 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases

- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.12.3	emr-5.12.2	emr-5.12.1	emr-5.12.0
AWS SDK for Java	1.11.267	1.11.267	1.11.267	1.11.267
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.0	1.4.0	1.4.0	1.4.0
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-

	emr-5.12.3	emr-5.12.2	emr-5.12.1	emr-5.12.0
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.0.0	1.0.0	1.0.0	1.0.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.188	0.188	0.188	0.188
Spark	2.2.1	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.12.3 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

> **Important**
>
> In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

5.12.3 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.9.0	Distributed copy application optimized for Amazon S3.
emrfs	2.21.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.4.0	Apache Flink command line client scripts and applications.

Component	Version	Description
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.3-amzn-0	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.3-amzn-0	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.3-amzn-0	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.3-amzn-0	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.3-amzn-0	MapReduce execution engine libraries for running a MapReduce application.

Component	Version	Description
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.3-amzn-0	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.0	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.0	Service for serving one or more HBase regions.
hbase-client	1.4.0	HBase command-line client.
hbase-rest-server	1.4.0	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.0	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.2-amzn-1	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.2-amzn-1	Service providing HCatalog, a table and storage management layer for distributed applications.

Component	Version	Description
hcatalog-webhcat-server	2.3.2-amzn-1	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.2-amzn-1	Hive command line client.
hive-hbase	2.3.2-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.2-amzn-1	Service for accepting Hive queries as web requests.
hue-server	4.1.0	Web application for analyzing data using Hadoop ecosystem applications
livy-server	0.4.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.
mxnet	1.0.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.1.85	Nvidia drivers and Cuda toolkit
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.

Component	Version	Description
phoenix-library	4.13.0-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.13.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.188	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.188	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
spark-client	2.2.1	Spark command-line clients.
spark-history-server	2.2.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.2.1	In-memory execution engine for YARN.
spark-yarn-slave	2.2.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.

Component	Version	Description
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.3	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.12.3 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.12.3 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.

Classifications	Description
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.

Classifications	Description
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.

Classifications	Description
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.

Classifications	Description
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.

Classifications	Description
presto-connector-postgresql	Change values in Presto's postgresql.properties file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.

Classifications	Description
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.12.2

5.12.2 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.12.2	emr-5.12.1	emr-5.12.0	emr-5.11.4
AWS SDK for Java	1.11.267	1.11.267	1.11.267	1.11.238
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4

	emr-5.12.2	emr-5.12.1	emr-5.12.0	emr-5.11.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.4.0	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.0	1.4.0	1.4.0	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.0.0	1.0.0	1.0.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.11.0

	emr-5.12.2	emr-5.12.1	emr-5.12.0	emr-5.11.4
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.188	0.188	0.188	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.12.2 release notes

The following release notes include information for Amazon EMR release 5.12.2. Changes are relative to 5.12.1.

Initial release date: August 29, 2018

Changes, enhancements, and resolved issues

- This release addresses a potential security vulnerability.

5.12.2 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.9.0	Distributed copy application optimized for Amazon S3.
emrfs	2.21.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.4.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.

Component	Version	Description
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.3-amzn-0	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.3-amzn-0	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.3-amzn-0	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.3-amzn-0	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.3-amzn-0	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN service for managing containers on an individual node.

Component	Version	Description
hadoop-yarn-resour cemanager	2.8.3-amzn-0	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.0	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.0	Service for serving one or more HBase regions.
hbase-client	1.4.0	HBase command-line client.
hbase-rest-server	1.4.0	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.0	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.2-amzn-1	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.2-amzn-1	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.2-amzn-1	HTTP endpoint providing a REST interface to HCatalog.

Component	Version	Description
hive-client	2.3.2-amzn-1	Hive command line client.
hive-hbase	2.3.2-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.2-amzn-1	Service for accepting Hive queries as web requests.
hue-server	4.1.0	Web application for analyzing data using Hadoop ecosystem applications
livy-server	0.4.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.
mxnet	1.0.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.1.85	Nvidia drivers and Cuda toolkit
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.13.0-HBase-1.4	The phoenix libraries for server and client

Component	Version	Description
phoenix-query-server	4.13.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.188	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.188	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
spark-client	2.2.1	Spark command-line clients.
spark-history-server	2.2.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.2.1	In-memory execution engine for YARN.
spark-yarn-slave	2.2.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.

Component	Version	Description
zeppelin-server	0.7.3	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.12.2 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see [Configure applications](#).

emr-5.12.2 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.

Classifications	Description
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.

Classifications	Description
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.

Classifications	Description
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.

Classifications	Description
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.

Classifications	Description
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.

Classifications	Description
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.12.1

5.12.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.12.1	emr-5.12.0	emr-5.11.4	emr-5.11.3
AWS SDK for Java	1.11.267	1.11.267	1.11.238	1.11.238
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-

	emr-5.12.1	emr-5.12.0	emr-5.11.4	emr-5.11.3
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.0	1.4.0	1.3.1	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.0.0	1.0.0	0.12.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.188	0.188	0.187	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1

	emr-5.12.1	emr-5.12.0	emr-5.11.4	emr-5.11.3
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.12.1 release notes

The following release notes include information for Amazon EMR release 5.12.1. Changes are relative to 5.12.0.

Initial release date: March 29, 2018

Changes, enhancements, and resolved issues

- Updated the Amazon Linux kernel of the defaultAmazon Linux AMI for Amazon EMR to address potential vulnerabilities.

5.12.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2

has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.9.0	Distributed copy application optimized for Amazon S3.
emrfs	2.21.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.4.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.

Component	Version	Description
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.3-amzn-0	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.3-amzn-0	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.3-amzn-0	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.3-amzn-0	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.3-amzn-0	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.3-amzn-0	YARN service for allocatin g and managing cluster resources and distributed applications.

Component	Version	Description
hadoop-yarn-timeline-server	2.8.3-amzn-0	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.0	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.0	Service for serving one or more HBase regions.
hbase-client	1.4.0	HBase command-line client.
hbase-rest-server	1.4.0	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.0	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.2-amzn-1	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.2-amzn-1	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.2-amzn-1	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.2-amzn-1	Hive command line client.
hive-hbase	2.3.2-amzn-1	Hive-hbase client.

Component	Version	Description
hive-metastore-server	2.3.2-amzn-1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.2-amzn-1	Service for accepting Hive queries as web requests.
hue-server	4.1.0	Web application for analyzing data using Hadoop ecosystem applications
livy-server	0.4.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.
mxnet	1.0.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.1.85	Nvidia drivers and Cuda toolkit
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.13.0-HBase-1.4	The phoenix libraries for server and client

Component	Version	Description
phoenix-query-server	4.13.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.188	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.188	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
spark-client	2.2.1	Spark command-line clients.
spark-history-server	2.2.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.2.1	In-memory execution engine for YARN.
spark-yarn-slave	2.2.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.

Component	Version	Description
zeppelin-server	0.7.3	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.12.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.12.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.

Classifications	Description
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.

Classifications	Description
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.

Classifications	Description
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.

Classifications	Description
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.

Classifications	Description
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.

Classifications	Description
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.12.0

5.12.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.12.0	emr-5.11.4	emr-5.11.3	emr-5.11.2
AWS SDK for Java	1.11.267	1.11.238	1.11.238	1.11.238
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-

	emr-5.12.0	emr-5.11.4	emr-5.11.3	emr-5.11.2
Delta	-	-	-	-
Flink	1.4.0	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.0	1.3.1	1.3.1	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.0.0	0.12.0	0.12.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.188	0.187	0.187	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1

	emr-5.12.0	emr-5.11.4	emr-5.11.3	emr-5.11.2
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.12.0 release notes

The following release notes include information for the Amazon EMR release 5.12.0. Changes are relative to 5.11.1.

Upgrades

- AWS SDK for Java 1.11.238 ⇒ 1.11.267. For more information, see the AWS SDK for Java Change Log on GitHub.
- Hadoop 2.7.3 ⇒ 2.8.3. For more information, see Apache Hadoop Releases.
- Flink 1.3.2 ⇒ 1.4.0. For more information, see the Apache Flink 1.4.0 Release Announcement.
- HBase 1.3.1 ⇒ 1.4.0. For more information, see the HBase Release Announcement.
- Hue 4.0.1 ⇒ 4.1.0. For more information, see the Release Notes.
- MxNet 0.12.0 ⇒ 1.0.0. For more information, see the MXNet Change Log on GitHub.
- Presto 0.187 ⇒ 0.188. For more information, see the Release Notes.

Changes, enhancements, and resolved issues

- **Hadoop**
  - The `yarn.resourcemanager.decommissioning.timeout` property has changed to `yarn.resourcemanager.nodemanager-graceful-decommission-timeout-secs`.

  You can use this property to customize cluster scale-down. For more information, see Cluster Scale-Down in the *Amazon EMR Management Guide*.

  - The Hadoop CLI added the `-d` option to the `cp` (copy) command, which specifies direct copy. You can use this to avoid creating an intermediary `.COPYING` file, which makes copying data between Amazon S3 faster. For more information, see HADOOP-12384.

- **Pig**
  - Added the `pig-env` configuration classification, which simplifies the configuration of Pig environment properties. For more information, see Configure applications.
- **Presto**
  - Added the `presto-connector-redshift` configuration classification, which you can use to configure values in the Presto `redshift.properties` configuration file. For more information, see Redshift Connector in Presto documentation, and Configure applications.
  - Presto support for EMRFS has been added and is the default configuration. Earlier Amazon EMR releases used PrestoS3FileSystem, which was the only option. For more information, see EMRFS and PrestoS3FileSystem configuration.

    Note

    If you query underlying data in Amazon S3 with Amazon EMR version 5.12.0, Presto errors can occur. This is because Presto fails to pick up configuration classification values from `emrfs-site.xml`. As a workaround, create an `emrfs` subdirectory under `usr/lib/presto/plugin/hive-hadoop2/` and create a symlink in `usr/lib/presto/plugin/hive-hadoop2/emrfs` to the existing `/usr/share/aws/emr/emrfs/conf/emrfs-site.xml` file. Then restart the presto-server process (`sudo presto-server stop` followed by `sudo presto-server start`).

- **Spark**
  - Backported SPARK-22036: BigDecimal multiplication sometimes returns null.

Known issues

- MXNet does not include OpenCV libraries.
- SparkR is not available for clusters created using a custom AMI because R is not installed by default on cluster nodes.

5.12.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.9.0	Distributed copy application optimized for Amazon S3.
emrfs	2.21.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.4.0	Apache Flink command line client scripts and applications.

Component	Version	Description
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.8.3-amzn-0	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.8.3-amzn-0	HDFS command-line client and library
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.8.3-amzn-0	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.8.3-amzn-0	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.8.3-amzn-0	MapReduce execution engine libraries for running a MapReduce application.

Component	Version	Description
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.8.3-amzn-0	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.4.0	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.4.0	Service for serving one or more HBase regions.
hbase-client	1.4.0	HBase command-line client.
hbase-rest-server	1.4.0	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.4.0	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.2-amzn-1	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.2-amzn-1	Service providing HCatalog, a table and storage management layer for distributed applications.

Component	Version	Description
hcatalog-webhcat-server	2.3.2-amzn-1	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.2-amzn-1	Hive command line client.
hive-hbase	2.3.2-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.2-amzn-1	Service for accepting Hive queries as web requests.
hue-server	4.1.0	Web application for analyzing data using Hadoop ecosystem applications
livy-server	0.4.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.
mxnet	1.0.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.1.85	Nvidia drivers and Cuda toolkit
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.

Component	Version	Description
phoenix-library	4.13.0-HBase-1.4	The phoenix libraries for server and client
phoenix-query-server	4.13.0-HBase-1.4	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.188	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.188	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
spark-client	2.2.1	Spark command-line clients.
spark-history-server	2.2.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.2.1	In-memory execution engine for YARN.
spark-yarn-slave	2.2.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.

Component	Version	Description
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.3	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.12.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.12.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.

Classifications	Description
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.

Classifications	Description
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.

Classifications	Description
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.

Classifications	Description
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-env	Change values in the Pig environment.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.

Classifications	Description
presto-connector-postgresql	Change values in Presto's postgresql.properties file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-redshift	Change values in Presto's redshift.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.

Classifications	Description
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.11.4

5.11.4 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.11.4	emr-5.11.3	emr-5.11.2	emr-5.11.1
AWS SDK for Java	1.11.238	1.11.238	1.11.238	1.11.238
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4

	emr-5.11.4	emr-5.11.3	emr-5.11.2	emr-5.11.1
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	0.12.0	0.12.0	0.12.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0

	emr-5.11.4	emr-5.11.3	emr-5.11.2	emr-5.11.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.187	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.11.4 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

> **Important**
>
> In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

5.11.4 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
aws-sagemaker-spark-sdk	1.0	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.8.0	Distributed copy application optimized for Amazon S3.
emrfs	2.20.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.3.2	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.

Component	Version	Description
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-6	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-6	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-6	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-6	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-6	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARN service for managing containers on an individual node.

Component	Version	Description
hadoop-yarn-resour cemanager	2.7.3-amzn-6	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.3.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.3.1	Service for serving one or more HBase regions.
hbase-client	1.3.1	HBase command-line client.
hbase-rest-server	1.3.1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.3.1	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.2-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.2-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.2-amzn-0	HTTP endpoint providing a REST interface to HCatalog.

Component	Version	Description
hive-client	2.3.2-amzn-0	Hive command line client.
hive-hbase	2.3.2-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.2-amzn-0	Service for accepting Hive queries as web requests.
hue-server	4.0.1	Web application for analyzing data using Hadoop ecosystem applications
livy-server	0.4.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.
mxnet	0.12.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.0.176	Nvidia drivers and Cuda toolkit
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.11.0-HBase-1.3	The phoenix libraries for server and client

Component	Version	Description
phoenix-query-server	4.11.0-HBase-1.3	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.187	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.187	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
spark-client	2.2.1	Spark command-line clients.
spark-history-server	2.2.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.2.1	In-memory execution engine for YARN.
spark-yarn-slave	2.2.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.

Component	Version	Description
zeppelin-server	0.7.3	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.11.4 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.11.4 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.

Classifications	Description
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.

Classifications	Description
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.

Classifications	Description
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.

Classifications	Description
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.

Classifications	Description
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.11.3

5.11.3 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.11.3	emr-5.11.2	emr-5.11.1	emr-5.11.0
AWS SDK for Java	1.11.238	1.11.238	1.11.238	1.11.238
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1

	emr-5.11.3	emr-5.11.2	emr-5.11.1	emr-5.11.0
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEn terpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	0.12.0	0.12.0	0.12.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.187	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.11.3	emr-5.11.2	emr-5.11.1	emr-5.11.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.11.3 release notes

The following release notes include information for Amazon EMR release 5.11.3. Changes are relative to 5.11.2.

Initial release date: July 18, 2019

Changes, enhancements, and resolved issues

- Updated the default Amazon Linux AMI for Amazon EMR to include important Linux kernel security updates, including the TCP SACK Denial of Service Issue (AWS-2019-005).

5.11.3 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.0	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.8.0	Distributed copy application optimized for Amazon S3.
emrfs	2.20.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.3.2	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.

Component	Version	Description
hadoop-client	2.7.3-amzn-6	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-6	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-6	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-6	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-6	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-6	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Service for retrieving current and historical information for YARN applications.

Component	Version	Description
hbase-hmaster	1.3.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.3.1	Service for serving one or more HBase regions.
hbase-client	1.3.1	HBase command-line client.
hbase-rest-server	1.3.1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.3.1	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.2-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.2-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.2-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.2-amzn-0	Hive command line client.
hive-hbase	2.3.2-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.

Component	Version	Description
hive-server2	2.3.2-amzn-0	Service for accepting Hive queries as web requests.
hue-server	4.0.1	Web application for analyzing data using Hadoop ecosystem applications
livy-server	0.4.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.
mxnet	0.12.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.0.176	Nvidia drivers and Cuda toolkit
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.11.0-HBase-1.3	The phoenix libraries for server and client
phoenix-query-server	4.11.0-HBase-1.3	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API

Component	Version	Description
presto-coordinator	0.187	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.187	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
spark-client	2.2.1	Spark command-line clients.
spark-history-server	2.2.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.2.1	In-memory execution engine for YARN.
spark-yarn-slave	2.2.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.3	Web-based notebook that enables interactive data analytics.

Component	Version	Description
zookeeper-server	3.4.10	Centralized service for maintaining configuration information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.11.3 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.11.3 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler.xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.

Classifications	Description
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.

Classifications	Description
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.

Classifications	Description
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.

Classifications	Description
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.

Classifications	Description
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.11.2

5.11.2 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.11.2	emr-5.11.1	emr-5.11.0	emr-5.10.1
AWS SDK for Java	1.11.238	1.11.238	1.11.238	1.11.221
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1

	emr-5.11.2	emr-5.11.1	emr-5.11.0	emr-5.10.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.1
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	0.12.0	0.12.0	0.12.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.187	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.11.2	emr-5.11.1	emr-5.11.0	emr-5.10.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.11.2 release notes

The following release notes include information for Amazon EMR release 5.11.2. Changes are relative to 5.11.1.

Initial release date: August 29, 2018

Changes, enhancements, and resolved issues

- This release addresses a potential security vulnerability.

5.11.2 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.0	Amazon SageMaker Spark SDK

Component	Version	Description
emr-ddb	4.5.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.8.0	Distributed copy application optimized for Amazon S3.
emrfs	2.20.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.3.2	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-6	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.

Component	Version	Description
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-6	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-6	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-6	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-6	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-6	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.3.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.

Component	Version	Description
hbase-region-server	1.3.1	Service for serving one or more HBase regions.
hbase-client	1.3.1	HBase command-line client.
hbase-rest-server	1.3.1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.3.1	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.2-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.2-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.2-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.2-amzn-0	Hive command line client.
hive-hbase	2.3.2-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.2-amzn-0	Service for accepting Hive queries as web requests.

Component	Version	Description
hue-server	4.0.1	Web application for analyzing data using Hadoop ecosystem applications
livy-server	0.4.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.
mxnet	0.12.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.0.176	Nvidia drivers and Cuda toolkit
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.11.0-HBase-1.3	The phoenix libraries for server and client
phoenix-query-server	4.11.0-HBase-1.3	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.187	Service for accepting queries and managing query execution among presto-wo rkers.

Component	Version	Description
presto-worker	0.187	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
spark-client	2.2.1	Spark command-line clients.
spark-history-server	2.2.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.2.1	In-memory execution engine for YARN.
spark-yarn-slave	2.2.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.3	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.

Component	Version	Description
zookeeper-client	3.4.10	ZooKeeper command line client.

5.11.2 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.11.2 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration

Classifications	Description
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.

Classifications	Description
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2- site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.

Classifications	Description
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.

Classifications	Description
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file

Classifications	Description
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.11.1

5.11.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- [Application versions in Amazon EMR 7.x releases]
- [Application versions in Amazon EMR 6.x releases]
- [Application versions in Amazon EMR 5.x releases]
- [Application versions in Amazon EMR 4.x releases]

Application version information

	emr-5.11.1	emr-5.11.0	emr-5.10.1	emr-5.10.0
AWS SDK for Java	1.11.238	1.11.238	1.11.221	1.11.221
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.2	2.3.2	2.3.1	2.3.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.1	2.3.1
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-

	emr-5.11.1	emr-5.11.0	emr-5.10.1	emr-5.10.0
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	0.12.0	0.12.0	0.12.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.187	0.187
Spark	2.2.1	2.2.1	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.11.1 release notes

The following release notes include information for the Amazon EMR 5.11.1 release. Changes are relative to the Amazon EMR 5.8.0 release.

Initial release date: January 22, 2018

Changes, enhancements, and resolved issues

- Updated the Amazon Linux kernel of the defaultAmazon Linux AMI for Amazon EMR to address vulnerabilities associated with speculative execution (CVE-2017-5715, CVE-2017-5753, and CVE-2017-5754). For more information, see https://aws.amazon.com/security/security-bulletins/AWS-2018-013/.

5.11.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
aws-sagemaker-spark-sdk	1.0	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.

Component	Version	Description
emr-s3-dist-cp	2.8.0	Distributed copy application optimized for Amazon S3.
emrfs	2.20.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.3.2	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-6	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-6	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFS service for tracking file names and block locations.

Component	Version	Description
hadoop-httpfs-server	2.7.3-amzn-6	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-6	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-6	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-6	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.3.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.3.1	Service for serving one or more HBase regions.
hbase-client	1.3.1	HBase command-line client.
hbase-rest-server	1.3.1	Service providing a RESTful HTTP endpoint for HBase.

Component	Version	Description
hbase-thrift-server	1.3.1	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.2-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.2-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.2-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.2-amzn-0	Hive command line client.
hive-hbase	2.3.2-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.2-amzn-0	Service for accepting Hive queries as web requests.
hue-server	4.0.1	Web application for analyzing data using Hadoop ecosystem applications
livy-server	0.4.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.

Component	Version	Description
mxnet	0.12.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.0.176	Nvidia drivers and Cuda toolkit
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.11.0-HBase-1.3	The phoenix libraries for server and client
phoenix-query-server	4.11.0-HBase-1.3	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.187	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.187	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
spark-client	2.2.1	Spark command-line clients.

Component	Version	Description
spark-history-server	2.2.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.2.1	In-memory execution engine for YARN.
spark-yarn-slave	2.2.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.3	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.11.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.11.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.

Classifications	Description
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.

Classifications	Description
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.

Classifications	Description
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.

Classifications	Description
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.

Classifications	Description
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.11.0

5.11.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases

- [Application versions in Amazon EMR 4.x releases]

Application version information

	emr-5.11.0	emr-5.10.1	emr-5.10.0	emr-5.9.1
AWS SDK for Java	1.11.238	1.11.221	1.11.221	1.11.183
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.2	2.3.1	2.3.1	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.1	2.3.1	2.3.0
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-

	emr-5.11.0	emr-5.10.1	emr-5.10.0	emr-5.9.1
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	0.12.0	0.12.0	0.12.0	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.187	0.184
Spark	2.2.1	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.11.0 release notes

The following release notes include information for the Amazon EMR release 5.11.0. Changes are relative to 5.10.0.

Upgrades

- Hive 2.3.2
- Spark 2.2.1

- SDK for Java 1.11.238

New features

- Spark
  - Added `spark.decommissioning.timeout.threshold` setting, which improves Spark decommissioning behavior when using Spot Instances. For more information, see Configuring node decommissioning behavior.
  - Added the `aws-sagemaker-spark-sdk` component to Spark, which installs Amazon SageMaker Spark and associated dependencies for Spark integration with Amazon SageMaker. You can use Amazon SageMaker Spark to construct Spark machine learning (ML) pipelines using Amazon SageMaker stages. For more information, see the SageMaker Spark Readme on GitHub and Using Apache Spark with Amazon SageMaker in the *Amazon SageMaker Developer Guide*.

Known issues

- MXNet does not include OpenCV libraries.
- Hive 2.3.2 sets `hive.compute.query.using.stats=true` by default. This causes queries to get data from existing statistics rather than directly from data, which could be confusing. For example, if you have a table with `hive.compute.query.using.stats=true` and upload new files to the table `LOCATION`, running a `SELECT COUNT(*)` query on the table returns the count from the statistics, rather than picking up the added rows.

  As a workaround, use the `ANALYZE TABLE` command to gather new statistics, or set `hive.compute.query.using.stats=false`. For more information, see Statistics in Hive in the Apache Hive documentation.

5.11.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
aws-sagemaker-spark-sdk	1.0	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.8.0	Distributed copy application optimized for Amazon S3.
emrfs	2.20.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.3.2	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.

Component	Version	Description
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-6	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-6	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-6	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-6	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-6	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARN service for managing containers on an individual node.

Component	Version	Description
hadoop-yarn-resour cemanager	2.7.3-amzn-6	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.3.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.3.1	Service for serving one or more HBase regions.
hbase-client	1.3.1	HBase command-line client.
hbase-rest-server	1.3.1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.3.1	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.2-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.2-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.2-amzn-0	HTTP endpoint providing a REST interface to HCatalog.

Component	Version	Description
hive-client	2.3.2-amzn-0	Hive command line client.
hive-hbase	2.3.2-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.2-amzn-0	Service for accepting Hive queries as web requests.
hue-server	4.0.1	Web application for analyzing data using Hadoop ecosystem applications
livy-server	0.4.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.
mxnet	0.12.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.0.176	Nvidia drivers and Cuda toolkit
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.11.0-HBase-1.3	The phoenix libraries for server and client

Component	Version	Description
phoenix-query-server	4.11.0-HBase-1.3	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.187	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.187	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
spark-client	2.2.1	Spark command-line clients.
spark-history-server	2.2.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.2.1	In-memory execution engine for YARN.
spark-yarn-slave	2.2.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.

Component	Version	Description
zeppelin-server	0.7.3	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.11.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.11.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.

Classifications	Description
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.

Classifications	Description
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.

Classifications	Description
hadoop-kms-env	Change values in the Hadoop KMS environment.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.

Classifications	Description
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.

Classifications	Description
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.10.1

5.10.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.10.1	emr-5.10.0	emr-5.9.1	emr-5.9.0
AWS SDK for Java	1.11.221	1.11.221	1.11.183	1.11.183
Python	2.7, 3.4	2.7, 3.4	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1

	emr-5.10.1	emr-5.10.0	emr-5.9.1	emr-5.9.0
HCatalog	2.3.1	2.3.1	2.3.0	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.1	2.3.1	2.3.0	2.3.0
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	0.12.0	0.12.0	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.184	0.184
Spark	2.2.0	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.10.1	emr-5.10.0	emr-5.9.1	emr-5.9.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.10.1 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

> **Important**
>
> In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

5.10.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.5.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.7.0	Distributed copy application optimized for Amazon S3.
emrfs	2.20.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.3.2	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-5	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.

Component	Version	Description
hadoop-hdfs-datanode	2.7.3-amzn-5	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-5	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-5	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-5	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-5	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-5	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-5	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-5	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-5	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.3.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.

Component	Version	Description
hbase-region-server	1.3.1	Service for serving one or more HBase regions.
hbase-client	1.3.1	HBase command-line client.
hbase-rest-server	1.3.1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.3.1	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.1-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.1-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.1-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.1-amzn-0	Hive command line client.
hive-hbase	2.3.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.1-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.1-amzn-0	Service for accepting Hive queries as web requests.

Component	Version	Description
hue-server	4.0.1	Web application for analyzing data using Hadoop ecosystem applications
livy-server	0.4.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.
mxnet	0.12.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.0.176	Nvidia drivers and Cuda toolkit
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.11.0-HBase-1.3	The phoenix libraries for server and client
phoenix-query-server	4.11.0-HBase-1.3	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.187	Service for accepting queries and managing query execution among presto-wo rkers.

Component	Version	Description
presto-worker	0.187	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
spark-client	2.2.0	Spark command-line clients.
spark-history-server	2.2.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.2.0	In-memory execution engine for YARN.
spark-yarn-slave	2.2.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.3	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.

Component	Version	Description
zookeeper-client	3.4.10	ZooKeeper command line client.

5.10.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.10.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration

Classifications	Description
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.

Classifications	Description
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2- site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.

Classifications	Description
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.

Classifications	Description
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file

Classifications	Description
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.10.0

5.10.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Livy, MXNet, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.10.0	emr-5.9.1	emr-5.9.0	emr-5.8.3
AWS SDK for Java	1.11.221	1.11.183	1.11.183	1.11.160
Python	2.7, 3.4	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.1	2.3.0	2.3.0	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.1	2.3.0	2.3.0	2.3.0
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	3.12.0
Iceberg	-	-	-	-

	emr-5.10.0	emr-5.9.1	emr-5.9.0	emr-5.8.3
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	-
MXNet	0.12.0	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.16.0
Presto	0.187	0.184	0.184	0.170
Spark	2.2.0	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.10.0 release notes

The following release notes include information for the Amazon EMR version 5.10.0 release. Changes are relative to the Amazon EMR 5.9.0 release.

Upgrades

- AWS SDK for Java 1.11.221
- Hive 2.3.1
- Presto 0.187

New features

- Added support for Kerberos authentication. For more information, see Use Kerberos Authentication in the *Amazon EMR Management Guide*
- Added support for IAM roles for EMRFS. For more information, see Configure IAM Roles for EMRFS Requests to Amazon S3 in the *Amazon EMR Management Guide*
- Added support for GPU-based P2 and P3 instance types. For more information, see Amazon EC2 P2 Instances and Amazon EC2 P3 Instances. NVIDIA driver 384.81 and CUDA driver 9.0.176 are installed on these instance types by default.
- Added support for Apache MXNet.

Changes, enhancements, and resolved issues

- Presto
  - Added support for using the AWS Glue Data Catalog as the default Hive metastore. For more information, see Using Presto with the AWS Glue Data Catalog.
  - Added support for geospatial functions.
  - Added spill to disk support for joins.
  - Added support for the Redshift connector.
- Spark
  - Backported SPARK-20640, which makes the rpc timeout and the retries for shuffle registration values configurable using `spark.shuffle.registration.timeout` and `spark.shuffle.registration.maxAttempts` properties.
  - Backported SPARK-21549, which corrects an error that occurs when writing custom OutputFormat to non-HDFS locations.
- Backported Hadoop-13270

- The Numpy, Scipy, and Matplotlib libraries have been removed from the base Amazon EMR AMI. If these libraries are required for your application, they are available in the application repository, so you can use a bootstrap action to install them on all nodes using `yum install`.
- The Amazon EMR base AMI no longer has application RPM packages included, so the RPM packages are no longer present on cluster nodes. Custom AMIs and the Amazon EMR base AMI now reference the RPM package repository in Amazon S3.
- Because of the introduction of per-second billing in Amazon EC2, the default **Scale down behavior** is now **Terminate at task completion** rather than **Terminate at instance hour**. For more information, see Configure Cluster Scale-Down.

Known issues

- MXNet does not include OpenCV libraries.
- Hive 2.3.1 sets `hive.compute.query.using.stats=true` by default. This causes queries to get data from existing statistics rather than directly from data, which could be confusing. For example, if you have a table with `hive.compute.query.using.stats=true` and upload new files to the table `LOCATION`, running a `SELECT COUNT(*)` query on the table returns the count from the statistics, rather than picking up the added rows.

  As a workaround, use the `ANALYZE TABLE` command to gather new statistics, or set `hive.compute.query.using.stats=false`. For more information, see Statistics in Hive in the Apache Hive documentation.

5.10.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.5.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.7.0	Distributed copy application optimized for Amazon S3.
emrfs	2.20.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.3.2	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-5	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.

Component	Version	Description
hadoop-hdfs-datanode	2.7.3-amzn-5	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-5	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-5	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-5	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-5	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-5	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-5	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-5	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-5	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.3.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.

Component	Version	Description
hbase-region-server	1.3.1	Service for serving one or more HBase regions.
hbase-client	1.3.1	HBase command-line client.
hbase-rest-server	1.3.1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.3.1	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.1-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.1-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.1-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.1-amzn-0	Hive command line client.
hive-hbase	2.3.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.1-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.1-amzn-0	Service for accepting Hive queries as web requests.

Component	Version	Description
hue-server	4.0.1	Web application for analyzing data using Hadoop ecosystem applications
livy-server	0.4.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.
mxnet	0.12.0	A flexible, scalable, and efficient library for deep learning.
mysql-server	5.5.54+	MySQL database server.
nvidia-cuda	9.0.176	Nvidia drivers and Cuda toolkit
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.11.0-HBase-1.3	The phoenix libraries for server and client
phoenix-query-server	4.11.0-HBase-1.3	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.187	Service for accepting queries and managing query execution among presto-wo rkers.

Component	Version	Description
presto-worker	0.187	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
spark-client	2.2.0	Spark command-line clients.
spark-history-server	2.2.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.2.0	In-memory execution engine for YARN.
spark-yarn-slave	2.2.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.3	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.

Component	Version	Description
zookeeper-client	3.4.10	ZooKeeper command line client.

5.10.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.10.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration

Classifications	Description
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.

Classifications	Description
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2- site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.

Classifications	Description
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.

Classifications	Description
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file

Classifications	Description
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.9.1

5.9.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Livy, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases

- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.9.1	emr-5.9.0	emr-5.8.3	emr-5.8.2
AWS SDK for Java	1.11.183	1.11.183	1.11.160	1.11.160
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.1	1.3.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.3.0	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.3.0	2.3.0
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-

	emr-5.9.1	emr-5.9.0	emr-5.8.3	emr-5.8.2
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.16.0	0.16.0
Presto	0.184	0.184	0.170	0.170
Spark	2.2.0	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.9.1 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

> **Important**
>
> In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

5.9.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.4.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.7.0	Distributed copy application optimized for Amazon S3.

Component	Version	Description
emrfs	2.19.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.3.2	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-4	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-4	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-4	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-4	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-4	HTTP endpoint for HDFS operations.

Component	Version	Description
hadoop-kms-server	2.7.3-amzn-4	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-4	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-4	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-4	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-4	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.3.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.3.1	Service for serving one or more HBase regions.
hbase-client	1.3.1	HBase command-line client.
hbase-rest-server	1.3.1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.3.1	Service providing a Thrift endpoint to HBase.

Component	Version	Description
hcatalog-client	2.3.0-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.0-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.0-amzn-0	Hive command line client.
hive-hbase	2.3.0-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.0-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.0-amzn-0	Service for accepting Hive queries as web requests.
hue-server	4.0.1	Web application for analyzing data using Hadoop ecosystem applications
livy-server	0.4.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.3.0	Oozie command-line client.

Component	Version	Description
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.11.0-HBase-1.3	The phoenix libraries for server and client
phoenix-query-server	4.11.0-HBase-1.3	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.184	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.184	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
spark-client	2.2.0	Spark command-line clients.
spark-history-server	2.2.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.2.0	In-memory execution engine for YARN.
spark-yarn-slave	2.2.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.

Component	Version	Description
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.2	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.9.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.9.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.

Classifications	Description
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.

Classifications	Description
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.

Classifications	Description
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.

Classifications	Description
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.

Classifications	Description
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.

Classifications	Description
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.9.0

5.9.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Livy, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.9.0	emr-5.8.3	emr-5.8.2	emr-5.8.1
AWS SDK for Java	1.11.183	1.11.160	1.11.160	1.11.160
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-

	emr-5.9.0	emr-5.8.3	emr-5.8.2	emr-5.8.1
Flink	1.3.2	1.3.1	1.3.1	1.3.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.3.0	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.3.0	2.3.0
Hudi	-	-	-	-
Hue	4.0.1	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.16.0	0.16.0	0.16.0
Presto	0.184	0.170	0.170	0.170
Spark	2.2.0	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6

	emr-5.9.0	emr-5.8.3	emr-5.8.2	emr-5.8.1
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.9.0 release notes

The following release notes include information for the Amazon EMR version 5.9.0 release. Changes are relative to the Amazon EMR 5.8.0 release.

Release date: October 5, 2017

Latest feature update: October 12, 2017

Upgrades

- AWS SDK for Java version 1.11.183
- Flink 1.3.2
- Hue 4.0.1
- Pig 0.17.0
- Presto 0.184

New features

- Added Livy support (version 0.4.0-incubating). For more information, see Apache Livy.
- Added support for Hue Notebook for Spark.
- Added support for i3-series Amazon EC2 instances (October 12, 2017).

Changes, enhancements, and resolved issues

- Spark
  - Added a new set of features that help ensure Spark handles node termination because of a manual resize or an automatic scaling policy request more gracefully. For more information, see Configuring node decommissioning behavior.
  - SSL is used instead of 3DES for in-transit encryption for the block transfer service, which enhances performance when using Amazon EC2 instance types with AES-NI.
  - Backported SPARK-21494.
- Zeppelin
  - Backported ZEPPELIN-2377.
- HBase
  - Added patch HBASE-18533, which allows additional values for HBase BucketCache configuration using the `hbase-site` configuration classification.
- Hue
  - Added AWS Glue Data Catalog support for the Hive query editor in Hue.
  - By default, superusers in Hue can access all files that Amazon EMR IAM roles are allowed to access. Newly created users do not automatically have permissions to access the Amazon S3 filebrowser and must have the `filebrowser.s3_access` permissions enabled for their group.
- Resolved an issue that caused underlying JSON data created using AWS Glue Data Catalog to be inaccessible.

Known issues

- Cluster launch fails when all applications are installed and the default Amazon EBS root volume size is not changed. As a workaround, use the `aws emr create-cluster` command from the AWS CLI and specify a larger `--ebs-root-volume-size` parameter.
- Hive 2.3.0 sets `hive.compute.query.using.stats=true` by default. This causes queries to get data from existing statistics rather than directly from data, which could be confusing. For example, if you have a table with `hive.compute.query.using.stats=true` and upload new files to the table `LOCATION`, running a `SELECT COUNT(*)` query on the table returns the count from the statistics, rather than picking up the added rows.

As a workaround, use the `ANALYZE TABLE` command to gather new statistics, or set `hive.compute.query.using.stats=false`. For more information, see Statistics in Hive in the Apache Hive documentation.

5.9.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.4.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.7.0	Distributed copy application optimized for Amazon S3.
emrfs	2.19.0	Amazon S3 connector for Hadoop ecosystem applicati ons.

Component	Version	Description
flink-client	1.3.2	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-4	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-4	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-4	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-4	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-4	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-4	Cryptographic key management server based on Hadoop's KeyProvider API.

Component	Version	Description
hadoop-mapred	2.7.3-amzn-4	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-4	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-4	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-4	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.3.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.3.1	Service for serving one or more HBase regions.
hbase-client	1.3.1	HBase command-line client.
hbase-rest-server	1.3.1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.3.1	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.0-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.

Component	Version	Description
hcatalog-server	2.3.0-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.0-amzn-0	Hive command line client.
hive-hbase	2.3.0-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.0-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.0-amzn-0	Service for accepting Hive queries as web requests.
hue-server	4.0.1	Web application for analyzing data using Hadoop ecosystem applications
livy-server	0.4.0-incubating	REST interface for interacting with Apache Spark
mahout-client	0.13.0	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.

Component	Version	Description
phoenix-library	4.11.0-HBase-1.3	The phoenix libraries for server and client
phoenix-query-server	4.11.0-HBase-1.3	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.184	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.184	Service for executing pieces of a query.
pig-client	0.17.0	Pig command-line client.
spark-client	2.2.0	Spark command-line clients.
spark-history-server	2.2.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.2.0	In-memory execution engine for YARN.
spark-yarn-slave	2.2.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.

Component	Version	Description
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.2	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.9.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.9.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.

Classifications	Description
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.

Classifications	Description
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.

Classifications	Description
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
livy-conf	Change values in Livy's livy.conf file.
livy-env	Change values in the Livy environment.
livy-log4j	Change Livy log4j.properties settings.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.

Classifications	Description
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.

Classifications	Description
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.

Classifications	Description
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.8.3

5.8.3 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.8.3	emr-5.8.2	emr-5.8.1	emr-5.8.0
AWS SDK for Java	1.11.160	1.11.160	1.11.160	1.11.160
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-

	emr-5.8.3	emr-5.8.2	emr-5.8.1	emr-5.8.0
Flink	1.3.1	1.3.1	1.3.1	1.3.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.3.0	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.3.0	2.3.0
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.170
Spark	2.2.0	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6

	emr-5.8.3	emr-5.8.2	emr-5.8.1	emr-5.8.0
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.8.3 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

> **Important**
>
> In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

5.8.3 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.4.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.6.0	Distributed copy application optimized for Amazon S3.
emrfs	2.18.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.3.1	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-3	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.

Component	Version	Description
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-3	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-3	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-3	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-3	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-3	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-3	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.3.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.

Component	Version	Description
hbase-region-server	1.3.1	Service for serving one or more HBase regions.
hbase-client	1.3.1	HBase command-line client.
hbase-rest-server	1.3.1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.3.1	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.0-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.0-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.0-amzn-0	Hive command line client.
hive-hbase	2.3.0-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.0-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.0-amzn-0	Service for accepting Hive queries as web requests.

Component	Version	Description
hue-server	3.12.0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.13.0	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.11.0-HBase-1.3	The phoenix libraries for server and client
phoenix-query-server	4.11.0-HBase-1.3	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.170	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.170	Service for executing pieces of a query.
pig-client	0.16.0-amzn-1	Pig command-line client.
spark-client	2.2.0	Spark command-line clients.
spark-history-server	2.2.0	Web UI for viewing logged events for the lifetime of a completed Spark application.

Component	Version	Description
spark-on-yarn	2.2.0	In-memory execution engine for YARN.
spark-yarn-slave	2.2.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.2	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.8.3 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.8.3 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.

Classifications	Description
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.

Classifications	Description
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.

Classifications	Description
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.

Classifications	Description
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.

Classifications	Description
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.8.2

5.8.2 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.8.2	emr-5.8.1	emr-5.8.0	emr-5.7.1
AWS SDK for Java	1.11.160	1.11.160	1.11.160	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8

	emr-5.8.2	emr-5.8.1	emr-5.8.0	emr-5.7.1
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.1	1.3.1	1.3.1	1.3.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.3.0	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.3.0	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0

	emr-5.8.2	emr-5.8.1	emr-5.8.0	emr-5.7.1
Presto	0.170	0.170	0.170	0.170
Spark	2.2.0	2.2.0	2.2.0	2.1.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.8.2 release notes

The following release notes include information for Amazon EMR release 5.8.2. Changes are relative to 5.8.1.

Initial release date: March 29, 2018

Changes, enhancements, and resolved issues

- Updated the Amazon Linux kernel of the defaultAmazon Linux AMI for Amazon EMR to address potential vulnerabilities.

5.8.2 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
emr-ddb	4.4.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.6.0	Distributed copy application optimized for Amazon S3.
emrfs	2.18.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.3.1	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.

Component	Version	Description
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-3	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-3	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-3	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-3	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-3	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARN service for managing containers on an individual node.
hadoop-yarn-resourcemanager	2.7.3-amzn-3	YARN service for allocating and managing cluster resources and distributed applications.

Component	Version	Description
hadoop-yarn-timeline-server	2.7.3-amzn-3	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.3.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.3.1	Service for serving one or more HBase regions.
hbase-client	1.3.1	HBase command-line client.
hbase-rest-server	1.3.1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.3.1	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.0-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.0-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.0-amzn-0	Hive command line client.
hive-hbase	2.3.0-amzn-0	Hive-hbase client.

Component	Version	Description
hive-metastore-server	2.3.0-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.0-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.12.0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.13.0	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.11.0-HBase-1.3	The phoenix libraries for server and client
phoenix-query-server	4.11.0-HBase-1.3	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.170	Service for accepting queries and managing query execution among presto-wo rkers.

Component	Version	Description
presto-worker	0.170	Service for executing pieces of a query.
pig-client	0.16.0-amzn-1	Pig command-line client.
spark-client	2.2.0	Spark command-line clients.
spark-history-server	2.2.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.2.0	In-memory execution engine for YARN.
spark-yarn-slave	2.2.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.2	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.

Component	Version	Description
zookeeper-client	3.4.10	ZooKeeper command line client.

5.8.2 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.8.2 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration

Classifications	Description
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.

Classifications	Description
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.

Classifications	Description
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.

Classifications	Description
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.

Classifications	Description
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.8.1

5.8.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.8.1	emr-5.8.0	emr-5.7.1	emr-5.7.0
AWS SDK for Java	1.11.160	1.11.160	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.1	1.3.1	1.3.0	1.3.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-5.8.1	emr-5.8.0	emr-5.7.1	emr-5.7.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.170
Spark	2.2.0	2.2.0	2.1.1	2.1.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.8.1 release notes

The following release notes include information for the Amazon EMR 5.8.1 release. Changes are relative to the Amazon EMR 5.8.0 release.

Initial release date: January 22, 2018

Changes, enhancements, and resolved issues

- Updated the Amazon Linux kernel of the defaultAmazon Linux AMI for Amazon EMR to address vulnerabilities associated with speculative execution (CVE-2017-5715, CVE-2017-5753, and CVE-2017-5754). For more information, see https://aws.amazon.com/security/security-bulletins/AWS-2018-013/.

5.8.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.4.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.6.0	Distributed copy application optimized for Amazon S3.
emrfs	2.18.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.3.1	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati

Component	Version	Description
		ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-3	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-3	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-3	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-3	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-3	MapReduce execution engine libraries for running a MapReduce application.

Component	Version	Description
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-3	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-3	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.3.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.3.1	Service for serving one or more HBase regions.
hbase-client	1.3.1	HBase command-line client.
hbase-rest-server	1.3.1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.3.1	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.0-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.0-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.

Component	Version	Description
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.0-amzn-0	Hive command line client.
hive-hbase	2.3.0-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.0-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.0-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.12.0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.13.0	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.11.0-HBase-1.3	The phoenix libraries for server and client
phoenix-query-server	4.11.0-HBase-1.3	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API

Component	Version	Description
presto-coordinator	0.170	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.170	Service for executing pieces of a query.
pig-client	0.16.0-amzn-1	Pig command-line client.
spark-client	2.2.0	Spark command-line clients.
spark-history-server	2.2.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.2.0	In-memory execution engine for YARN.
spark-yarn-slave	2.2.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.2	Web-based notebook that enables interactive data analytics.

Component	Version	Description
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.8.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.8.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.

Classifications	Description
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.

Classifications	Description
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.

Classifications	Description
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.

Classifications	Description
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.

Classifications	Description
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.8.0

5.8.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.8.0	emr-5.7.1	emr-5.7.0	emr-5.6.1
AWS SDK for Java	1.11.160	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.1	1.3.0	1.3.0	1.2.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.0
HCatalog	2.3.0	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0

	emr-5.8.0	emr-5.7.1	emr-5.7.0	emr-5.6.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.9.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.170
Spark	2.2.0	2.1.1	2.1.1	2.1.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.8.0 release notes

The following release notes include information for the Amazon EMR version 5.8.0 release. Changes are relative to the Amazon EMR 5.7.0 release.

Initial release date: August 10, 2017

Latest feature update: September 25, 2017

Upgrades

- AWS SDK 1.11.160
- Flink 1.3.1
- Hive 2.3.0. For more information, see Release Notes on the Apache Hive site.
- Spark 2.2.0. For more information, see Release Notes on the Apache Spark site.

New features

- Added support for viewing application history (September 25, 2017). For more information, see Viewing Application History in the *Amazon EMR Management Guide*.

Changes, enhancements, and resolved issues

- **Integration with AWS Glue Data Catalog**
  - Added ability for Hive and Spark SQL to use AWS Glue Data Catalog as the Hive metadata store. For more information, see Using the AWS Glue Data Catalog as the metastore for Hive and Use the AWS Glue Data Catalog as the metastore for Spark SQL.
- Added **Application history** to cluster details, which allows you to view historical data for YARN applications and additional details for Spark applications. For more information, see View Application History in the *Amazon EMR Management Guide*.
- **Oozie**
  - Backported OOZIE-2748.
- **Hue**
  - Backported HUE-5859
- **HBase**

- Added patch to expose the HBase master server start time through Java Management Extensions (JMX) using `getMasterInitializedTime`.
- Added patch that improves cluster start time.

Known issues

- Cluster launch fails when all applications are installed and the default Amazon EBS root volume size is not changed. As a workaround, use the `aws emr create-cluster` command from the AWS CLI and specify a larger `--ebs-root-volume-size` parameter.
- Hive 2.3.0 sets `hive.compute.query.using.stats=true` by default. This causes queries to get data from existing statistics rather than directly from data, which could be confusing. For example, if you have a table with `hive.compute.query.using.stats=true` and upload new files to the table `LOCATION`, running a `SELECT COUNT(*)` query on the table returns the count from the statistics, rather than picking up the added rows.

  As a workaround, use the `ANALYZE TABLE` command to gather new statistics, or set `hive.compute.query.using.stats=false`. For more information, see Statistics in Hive in the Apache Hive documentation.
- **Spark**—When using Spark, there is a file handler leak issue with the apppusher daemon, which can appear for a long-running Spark job after several hours or days. To fix the issue, connect to the primary node and type `sudo /etc/init.d/apppusher stop`. This stops that apppusher daemon, which Amazon EMR will restart automatically.
- **Application history**
  - Historical data for dead Spark executors is not available.
  - Application history is not available for clusters that use a security configuration to enable in-flight encryption.

5.8.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
emr-ddb	4.4.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.4.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.4.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.6.0	Distributed copy application optimized for Amazon S3.
emrfs	2.18.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.3.1	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.

Component	Version	Description
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-3	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-3	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-3	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-3	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-3	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-3	YARN service for allocatin g and managing cluster resources and distributed applications.

Component	Version	Description
hadoop-yarn-timeline-server	2.7.3-amzn-3	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.3.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.3.1	Service for serving one or more HBase regions.
hbase-client	1.3.1	HBase command-line client.
hbase-rest-server	1.3.1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.3.1	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.3.0-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.3.0-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.3.0-amzn-0	Hive command line client.
hive-hbase	2.3.0-amzn-0	Hive-hbase client.

Component	Version	Description
hive-metastore-server	2.3.0-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.3.0-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.12.0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.13.0	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.11.0-HBase-1.3	The phoenix libraries for server and client
phoenix-query-server	4.11.0-HBase-1.3	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.170	Service for accepting queries and managing query execution among presto-wo rkers.

Component	Version	Description
presto-worker	0.170	Service for executing pieces of a query.
pig-client	0.16.0-amzn-1	Pig command-line client.
spark-client	2.2.0	Spark command-line clients.
spark-history-server	2.2.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.2.0	In-memory execution engine for YARN.
spark-yarn-slave	2.2.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.2	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.

Component	Version	Description
zookeeper-client	3.4.10	ZooKeeper command line client.

5.8.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see [Configure applications](#).

emr-5.8.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration

Classifications	Description
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.

Classifications	Description
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.

Classifications	Description
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.

Classifications	Description
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.

Classifications	Description
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.7.1

5.7.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.7.1	emr-5.7.0	emr-5.6.1	emr-5.6.0
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.0	1.3.0	1.2.1	1.2.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-5.7.1	emr-5.7.0	emr-5.6.1	emr-5.6.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.9.0	4.9.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.1	2.1.1	2.1.1	2.1.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.7.1 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

> **Important**
>
> In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

5.7.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.3.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.3.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.3.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.5.0	Distributed copy application optimized for Amazon S3.
emrfs	2.18.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.3.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati

Component	Version	Description
		ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-2	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-2	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-2	MapReduce execution engine libraries for running a MapReduce application.

Component	Version	Description
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-2	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.3.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.3.1	Service for serving one or more HBase regions.
hbase-client	1.3.1	HBase command-line client.
hbase-rest-server	1.3.1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.3.1	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.1.1-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.

Component	Version	Description
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.1.1-amzn-0	Hive command line client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.1.1-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.12.0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.13.0	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.11.0-HBase-1.3	The phoenix libraries for server and client
phoenix-query-server	4.11.0-HBase-1.3	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API

Component	Version	Description
presto-coordinator	0.170	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.170	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.1.1	Spark command-line clients.
spark-history-server	2.1.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.1.1	In-memory execution engine for YARN.
spark-yarn-slave	2.1.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.2	Web-based notebook that enables interactive data analytics.

Component	Version	Description
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.7.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.7.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.

Classifications	Description
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.

Classifications	Description
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.

Classifications	Description
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.

Classifications	Description
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.

Classifications	Description
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.7.0

5.7.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.7.0	emr-5.6.1	emr-5.6.0	emr-5.5.4
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.0	1.2.1	1.2.1	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0

	emr-5.7.0	emr-5.6.1	emr-5.6.0	emr-5.5.4
Iceberg	-	-	-	-
JupyterEn terpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.1	2.1.1	2.1.1	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQ L)	-	-	-	-
Zeppelin	0.7.2	0.7.1	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.7.0 release notes

The following release notes include information for the Amazon EMR 5.7.0 release. Changes are relative to the Amazon EMR 5.6.0 release.

Release date: July 13, 2017

Upgrades

- Flink 1.3.0
- Phoenix 4.11.0
- Zeppelin 0.7.2

New features

- Added the ability to specify a customAmazon Linux AMI when you create a cluster. For more information, see Using a Custom AMI.

Changes, enhancements, and resolved issues

- **HBase**
  - Added capability to configure HBase read-replica clusters. See Using a Read-Replica Cluster.
  - Multiple bug fixes and enhancements
- **Presto** - added ability to configure `node.properties`.
- **YARN** - added ability to configure `container-log4j.properties`
- **Sqoop** - backported SQOOP-2880, which introduces an argument that allows you to set the Sqoop temporary directory.

5.7.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
emr-ddb	4.3.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.3.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.3.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.5.0	Distributed copy application optimized for Amazon S3.
emrfs	2.18.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.3.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.

Component	Version	Description
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-2	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-2	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-2	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-2	YARN service for allocatin g and managing cluster resources and distributed applications.

Component	Version	Description
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.3.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.3.1	Service for serving one or more HBase regions.
hbase-client	1.3.1	HBase command-line client.
hbase-rest-server	1.3.1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.3.1	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.1.1-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.1.1-amzn-0	Hive command line client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.

Component	Version	Description
hive-metastore-server	2.1.1-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.1.1-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.12.0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.13.0	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.11.0-HBase-1.3	The phoenix libraries for server and client
phoenix-query-server	4.11.0-HBase-1.3	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.170	Service for accepting queries and managing query execution among presto-wo rkers.

Component	Version	Description
presto-worker	0.170	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.1.1	Spark command-line clients.
spark-history-server	2.1.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.1.1	In-memory execution engine for YARN.
spark-yarn-slave	2.1.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.2	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.

Component	Version	Description
zookeeper-client	3.4.10	ZooKeeper command line client.

5.7.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see [Configure applications](#).

emr-5.7.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration

Classifications	Description
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.properties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hbase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.properties file.

Classifications	Description
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.

Classifications	Description
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.

Classifications	Description
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.

Classifications	Description
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.6.1

5.6.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.6.1	emr-5.6.0	emr-5.5.4	emr-5.5.3
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.1	1.2.1	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-5.6.1	emr-5.6.0	emr-5.5.4	emr-5.5.3
Mahout	0.13.0	0.13.0	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.1	2.1.1	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.6.1 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

Important

In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

5.6.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.3.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.3.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.3.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.5.0	Distributed copy application optimized for Amazon S3.
emrfs	2.17.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.2.1	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati

Component	Version	Description
		ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-2	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-2	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-2	MapReduce execution engine libraries for running a MapReduce application.

Component	Version	Description
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-2	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.3.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.3.1	Service for serving one or more HBase regions.
hbase-client	1.3.1	HBase command-line client.
hbase-rest-server	1.3.1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.3.1	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.1.1-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.

Component	Version	Description
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.1.1-amzn-0	Hive command line client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.1.1-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.12.0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.13.0	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.9.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.9.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API

Component	Version	Description
presto-coordinator	0.170	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.170	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.1.1	Spark command-line clients.
spark-history-server	2.1.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.1.1	In-memory execution engine for YARN.
spark-yarn-slave	2.1.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.1	Web-based notebook that enables interactive data analytics.

Component	Version	Description
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.6.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.6.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.

Classifications	Description
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.

Classifications	Description
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.

Classifications	Description
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.

Classifications	Description
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.

Classifications	Description
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.6.0

5.6.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.6.0	emr-5.5.4	emr-5.5.3	emr-5.5.2
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.1	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0

	emr-5.6.0	emr-5.5.4	emr-5.5.3	emr-5.5.2
Iceberg	-	-	-	-
JupyterEn terpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.1	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQ L)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.6.0 release notes

The following release notes include information for the Amazon EMR 5.6.0 release. Changes are relative to the Amazon EMR 5.5.0 release.

Release date: June 5, 2017

Upgrades

- Flink 1.2.1
- HBase 1.3.1
- Mahout 0.13.0. This is the first version of Mahout to support Spark 2.x in Amazon EMR version 5.0 and later.
- Spark 2.1.1

Changes, enhancements, and resolved issues

- **Presto**
  - Added the ability to enable SSL/TLS secured communication between Presto nodes by enabling in-transit encryption using a security configuration. For more information, see In-transit Data Encryption.
  - Backported Presto 7661, which adds the `VERBOSE` option to the `EXPLAIN ANALYZE` statement to report more detailed, low-level statistics about a query plan.

5.6.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.3.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.3.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.3.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.5.0	Distributed copy application optimized for Amazon S3.
emrfs	2.17.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.2.1	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-2	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.

Component	Version	Description
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-2	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-2	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-2	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.3.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.

Component	Version	Description
hbase-region-server	1.3.1	Service for serving one or more HBase regions.
hbase-client	1.3.1	HBase command-line client.
hbase-rest-server	1.3.1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.3.1	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.1.1-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.1.1-amzn-0	Hive command line client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.1.1-amzn-0	Service for accepting Hive queries as web requests.

Component	Version	Description
hue-server	3.12.0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.13.0	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.9.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.9.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.170	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.170	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.1.1	Spark command-line clients.
spark-history-server	2.1.1	Web UI for viewing logged events for the lifetime of a completed Spark application.

Component	Version	Description
spark-on-yarn	2.1.1	In-memory execution engine for YARN.
spark-yarn-slave	2.1.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.6.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.6.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb ase.properties file.

Classifications	Description
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.

Classifications	Description
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.

Classifications	Description
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-env	Change values in Presto's presto-env.sh file.
presto-node	Change values in Presto's node.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.

Classifications	Description
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.

Classifications	Description
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.5.4

5.5.4 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.5.4	emr-5.5.3	emr-5.5.2	emr-5.5.1
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8

	emr-5.5.4	emr-5.5.3	emr-5.5.2	emr-5.5.1
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0

	emr-5.5.4	emr-5.5.3	emr-5.5.2	emr-5.5.1
Presto	0.170	0.170	0.170	0.170
Spark	2.1.0	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.5.4 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

> **Important**
>
> In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

5.5.4 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0.

For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.3.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.3.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.3.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.5.0	Distributed copy application optimized for Amazon S3.
emrfs	2.16.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.2.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.

Component	Version	Description
hadoop-client	2.7.3-amzn-2	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-2	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-2	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-2	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service for retrieving current and historical information for YARN applications.

Component	Version	Description
hbase-hmaster	1.3.0	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.3.0	Service for serving one or more HBase regions.
hbase-client	1.3.0	HBase command-line client.
hbase-rest-server	1.3.0	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.3.0	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.1.1-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.1.1-amzn-0	Hive command line client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.

Component	Version	Description
hive-server2	2.1.1-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.12.0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.9.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.9.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.170	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.170	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.1.0	Spark command-line clients.

Component	Version	Description
spark-history-server	2.1.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.1.0	In-memory execution engine for YARN.
spark-yarn-slave	2.1.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.5.4 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.5.4 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.

Classifications	Description
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.

Classifications	Description
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.

Classifications	Description
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.

Classifications	Description
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.

Classifications	Description
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.5.3

5.5.3 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.5.3	emr-5.5.2	emr-5.5.1	emr-5.5.0
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75

	emr-5.5.3	emr-5.5.2	emr-5.5.1	emr-5.5.0
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0

	emr-5.5.3	emr-5.5.2	emr-5.5.1	emr-5.5.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.0	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.5.3 release notes

The following release notes include information for Amazon EMR release 5.5.3. Changes are relative to 5.5.2.

Initial release date: August 29, 2018

Changes, enhancements, and resolved issues

- This release addresses a potential security vulnerability.

5.5.3 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the

most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
emr-ddb	4.3.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.3.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.3.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.5.0	Distributed copy application optimized for Amazon S3.
emrfs	2.16.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.2.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.

Component	Version	Description
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-2	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-2	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-2	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN service for managing containers on an individual node.

Component	Version	Description
hadoop-yarn-resour cemanager	2.7.3-amzn-2	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.3.0	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.3.0	Service for serving one or more HBase regions.
hbase-client	1.3.0	HBase command-line client.
hbase-rest-server	1.3.0	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.3.0	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.1.1-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP endpoint providing a REST interface to HCatalog.

Component	Version	Description
hive-client	2.1.1-amzn-0	Hive command line client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.1.1-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.12.0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.9.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.9.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API

Component	Version	Description
presto-coordinator	0.170	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.170	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.1.0	Spark command-line clients.
spark-history-server	2.1.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.1.0	In-memory execution engine for YARN.
spark-yarn-slave	2.1.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.1	Web-based notebook that enables interactive data analytics.

Component	Version	Description
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.5.3 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.5.3 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.

Classifications	Description
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.

Classifications	Description
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.

Classifications	Description
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.

Classifications	Description
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file

Classifications	Description
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.5.2

5.5.2 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases

- [Application versions in Amazon EMR 6.x releases]
- [Application versions in Amazon EMR 5.x releases]
- [Application versions in Amazon EMR 4.x releases]

Application version information

	emr-5.5.2	emr-5.5.1	emr-5.5.0	emr-5.4.1
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-

	emr-5.5.2	emr-5.5.1	emr-5.5.0	emr-5.4.1
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.166
Spark	2.1.0	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.0
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.9

5.5.2 release notes

The following release notes include information for Amazon EMR release 5.5.2. Changes are relative to 5.5.1.

Initial release date: March 29, 2018

Changes, enhancements, and resolved issues

- Updated the Amazon Linux kernel of the defaultAmazon Linux AMI for Amazon EMR to address potential vulnerabilities.

5.5.2 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.3.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.3.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.3.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.5.0	Distributed copy application optimized for Amazon S3.
emrfs	2.16.0	Amazon S3 connector for Hadoop ecosystem applicati ons.

Component	Version	Description
flink-client	1.2.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-2	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-2	Cryptographic key management server based on Hadoop's KeyProvider API.

Component	Version	Description
hadoop-mapred	2.7.3-amzn-2	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-2	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.3.0	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.3.0	Service for serving one or more HBase regions.
hbase-client	1.3.0	HBase command-line client.
hbase-rest-server	1.3.0	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.3.0	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.1.1-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.

Component	Version	Description
hcatalog-server	2.1.1-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.1.1-amzn-0	Hive command line client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.1.1-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.12.0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.9.0-HBase-1.2	The phoenix libraries for server and client

Component	Version	Description
phoenix-query-server	4.9.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.170	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.170	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.1.0	Spark command-line clients.
spark-history-server	2.1.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.1.0	In-memory execution engine for YARN.
spark-yarn-slave	2.1.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.

Component	Version	Description
zeppelin-server	0.7.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.5.2 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.5.2 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.

Classifications	Description
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.

Classifications	Description
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.

Classifications	Description
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.

Classifications	Description
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.

Classifications	Description
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.5.1

5.5.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.5.1	emr-5.5.0	emr-5.4.1	emr-5.4.0
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.11.0	3.11.0

	emr-5.5.1	emr-5.5.0	emr-5.4.1	emr-5.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.166	0.166
Spark	2.1.0	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.0	0.7.0
ZooKeeper	3.4.10	3.4.10	3.4.9	3.4.9

5.5.1 release notes

The following release notes include information for the Amazon EMR 5.5.1 release. Changes are relative to the Amazon EMR 5.5.0 release.

Initial release date: January 22, 2018

Changes, enhancements, and resolved issues

- Updated the Amazon Linux kernel of the defaultAmazon Linux AMI for Amazon EMR to address vulnerabilities associated with speculative execution (CVE-2017-5715, CVE-2017-5753, and CVE-2017-5754). For more information, see https://aws.amazon.com/security/security-bulletins/AWS-2018-013/.

5.5.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.3.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.3.0	Extra convenience libraries for the Hadoop ecosystem.

Component	Version	Description
emr-kinesis	3.3.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.5.0	Distributed copy application optimized for Amazon S3.
emrfs	2.16.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.2.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-2	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS command-line client and library

Component	Version	Description
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-2	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-2	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-2	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.3.0	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.3.0	Service for serving one or more HBase regions.
hbase-client	1.3.0	HBase command-line client.

Component	Version	Description
hbase-rest-server	1.3.0	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.3.0	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.1.1-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.1.1-amzn-0	Hive command line client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.1.1-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.12.0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.

Component	Version	Description
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.9.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.9.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.170	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.170	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.1.0	Spark command-line clients.
spark-history-server	2.1.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.1.0	In-memory execution engine for YARN.
spark-yarn-slave	2.1.0	Apache Spark libraries needed by YARN slaves.

Component	Version	Description
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.5.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.5.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.

Classifications	Description
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.

Classifications	Description
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file

Classifications	Description
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.

Classifications	Description
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.

Classifications	Description
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.5.0

5.5.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.5.0	emr-5.4.1	emr-5.4.0	emr-5.3.2
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.1.4
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.2.3

	emr-5.5.0	emr-5.4.1	emr-5.4.0	emr-5.3.2
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.11.0	3.11.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.166	0.166	0.157.1
Spark	2.1.0	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.5.0	emr-5.4.1	emr-5.4.0	emr-5.3.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.0	0.7.0	0.6.2
ZooKeeper	3.4.10	3.4.9	3.4.9	3.4.9

5.5.0 release notes

The following release notes include information for the Amazon EMR 5.5.0 release. Changes are relative to the Amazon EMR 5.4.0 release.

Release date: April 26, 2017

Upgrades

- Hue 3.12
- Presto 0.170
- Zeppelin 0.7.1
- ZooKeeper 3.4.10

Changes, enhancements, and resolved issues

- **Spark**
  - Backported Spark Patch (SPARK-20115) Fix DAGScheduler to recompute all the lost shuffle blocks when external shuffle service is unavailable to version 2.1.0 of Spark, which is included in this release.
- **Flink**
  - Flink is now built with Scala 2.11. If you use the Scala API and libraries, we recommend that you use Scala 2.11 in your projects.
  - Addressed an issue where `HADOOP_CONF_DIR` and `YARN_CONF_DIR` defaults were not properly set, so `start-scala-shell.sh` failed to work. Also added the ability to set these values using `env.hadoop.conf.dir` and `env.yarn.conf.dir` in `/etc/flink/conf/flink-conf.yaml` or the `flink-conf` configuration classification.

  - Introduced a new EMR-specific command, `flink-scala-shell` as a wrapper for `start-scala-shell.sh`. We recommend using this command instead of `start-scala-shell`. The new command simplifies execution. For example, `flink-scala-shell -n 2` starts a Flink Scala shell with a task parallelism of 2.
  - Introduced a new EMR-specific command, `flink-yarn-session` as a wrapper for `yarn-session.sh`. We recommend using this command instead of `yarn-session`. The new command simplifies execution. For example, `flink-yarn-session -d -n 2` starts a long-running Flink session in a detached state with two task managers.
  - Addressed [(FLINK-6125) Commons httpclient is not shaded anymore in Flink 1.2](#).
- **Presto**
  - Added support for LDAP authentication. Using LDAP with Presto on Amazon EMR requires that you enable HTTPS access for the Presto coordinator (`http-server.https.enabled=true` in `config.properties`). For configuration details, see [LDAP Authentication](#) in Presto documentation.
  - Added support for `SHOW GRANTS`.
- **Amazon EMR Base Linux AMI**
  - Amazon EMR releases are now based on Amazon Linux 2017.03. For more information, see [Amazon Linux AMI 2017.03 Release Notes](#).
  - Removed Python 2.6 from the Amazon EMR base Linux image. Python 2.7 and 3.4 are installed by default. You can install Python 2.6 manually if necessary.

5.5.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.3.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.3.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.3.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.5.0	Distributed copy application optimized for Amazon S3.
emrfs	2.16.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.2.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-2	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.

Component	Version	Description
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-2	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-2	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-2	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.3.0	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.

Component	Version	Description
hbase-region-server	1.3.0	Service for serving one or more HBase regions.
hbase-client	1.3.0	HBase command-line client.
hbase-rest-server	1.3.0	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.3.0	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.1.1-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.1.1-amzn-0	Hive command line client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.1.1-amzn-0	Service for accepting Hive queries as web requests.

Component	Version	Description
hue-server	3.12.0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.9.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.9.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.170	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.170	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.1.0	Spark command-line clients.
spark-history-server	2.1.0	Web UI for viewing logged events for the lifetime of a completed Spark application.

Component	Version	Description
spark-on-yarn	2.1.0	In-memory execution engine for YARN.
spark-yarn-slave	2.1.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.10	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.10	ZooKeeper command line client.

5.5.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.5.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.

Classifications	Description
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.

Classifications	Description
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environment.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.

Classifications	Description
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.

Classifications	Description
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.

Classifications	Description
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.4.1

5.4.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.4.1	emr-5.4.0	emr-5.3.2	emr-5.3.1
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-

	emr-5.4.1	emr-5.4.0	emr-5.3.2	emr-5.3.1
Flink	1.2.0	1.2.0	1.1.4	1.1.4
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.2.3	1.2.3
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.11.0	3.11.0	3.11.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.166	0.166	0.157.1	0.157.1
Spark	2.1.0	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6

	emr-5.4.1	emr-5.4.0	emr-5.3.2	emr-5.3.1
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.0	0.7.0	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.9

5.4.1 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

> **Important**
>
> In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

5.4.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.2.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.3.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.15.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.2.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.

Component	Version	Description
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.3.0	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.

Component	Version	Description
hbase-region-server	1.3.0	Service for serving one or more HBase regions.
hbase-client	1.3.0	HBase command-line client.
hbase-rest-server	1.3.0	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.3.0	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.1.1-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.1.1-amzn-0	Hive command line client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.1.1-amzn-0	Service for accepting Hive queries as web requests.

Component	Version	Description
hue-server	3.11.0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.9.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.9.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.166	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.166	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.1.0	Spark command-line clients.
spark-history-server	2.1.0	Web UI for viewing logged events for the lifetime of a completed Spark application.

Component	Version	Description
spark-on-yarn	2.1.0	In-memory execution engine for YARN.
spark-yarn-slave	2.1.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.0	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.9	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.9	ZooKeeper command line client.

5.4.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.4.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.

Classifications	Description
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.

Classifications	Description
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environment.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.

Classifications	Description
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.

Classifications	Description
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.

Classifications	Description
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.4.0

5.4.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.4.0	emr-5.3.2	emr-5.3.1	emr-5.3.0
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-

	emr-5.4.0	emr-5.3.2	emr-5.3.1	emr-5.3.0
Flink	1.2.0	1.1.4	1.1.4	1.1.4
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.2.3	1.2.3	1.2.3
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.11.0	3.11.0	3.11.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.166	0.157.1	0.157.1	0.157.1
Spark	2.1.0	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6

	emr-5.4.0	emr-5.3.2	emr-5.3.1	emr-5.3.0
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.0	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.9

5.4.0 release notes

The following release notes include information for the Amazon EMR 5.4.0 release. Changes are relative to the Amazon EMR 5.3.0 release.

Release date: March 08, 2017

Upgrades

- Upgraded to Flink 1.2.0
- Upgraded to HBase 1.3.0
- Upgraded to Phoenix 4.9.0

> **Note**
>
> If you upgrade from an earlier version of Amazon EMR to Amazon EMR version 5.4.0 or later and use secondary indexing, upgrade local indexes as described in the [Apache Phoenix documentation](). Amazon EMR removes the required configurations from the `hbase-site` classification, but indexes need to be repopulated. Online and offline upgrade of indexes are supported. Online upgrades are the default, which means indexes are repopulated while initializing from Phoenix clients of version 4.8.0 or greater. To specify offline upgrades, set the `phoenix.client.localIndexUpgrade` configuration to false in the `phoenix-site` classification, and then SSH to the primary node to run `psql [zookeeper] -1`.

- Upgraded to Presto 0.166

- Upgraded to Zeppelin 0.7.0

Changes and enhancements

- Added support for r4 instances. See Amazon EC2 Instance Types.

5.4.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.2.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.3.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.

Component	Version	Description
emrfs	2.15.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.2.0	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-1	HTTP endpoint for HDFS operations.

Component	Version	Description
hadoop-kms-server	2.7.3-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.3.0	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.3.0	Service for serving one or more HBase regions.
hbase-client	1.3.0	HBase command-line client.
hbase-rest-server	1.3.0	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.3.0	Service providing a Thrift endpoint to HBase.

Component	Version	Description
hcatalog-client	2.1.1-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.1.1-amzn-0	Hive command line client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server2	2.1.1-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.11.0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.

Component	Version	Description
phoenix-library	4.9.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.9.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.166	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.166	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.1.0	Spark command-line clients.
spark-history-server	2.1.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.1.0	In-memory execution engine for YARN.
spark-yarn-slave	2.1.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.

Component	Version	Description
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.7.0	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.9	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.9	ZooKeeper command line client.

5.4.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.4.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.

Classifications	Description
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.

Classifications	Description
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.

Classifications	Description
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.

Classifications	Description
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.

Classifications	Description
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.3.2

5.3.2 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.3.2	emr-5.3.1	emr-5.3.0	emr-5.2.3
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.4	1.1.4	1.1.4	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.1	2.1.1	2.1.1	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.0
Hudi	-	-	-	-

	emr-5.3.2	emr-5.3.1	emr-5.3.0	emr-5.2.3
Hue	3.11.0	3.11.0	3.11.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.157.1	0.157.1	0.157.1	0.157.1
Spark	2.1.0	2.1.0	2.1.0	2.0.2
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.9

5.3.2 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

> **Important**
>
> In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

5.3.2 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.2.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.2.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.

Component	Version	Description
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.14.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.1.4	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS service for tracking file names and block locations.

Component	Version	Description
hadoop-httpfs-server	2.7.3-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.3	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.3	Service for serving one or more HBase regions.
hbase-client	1.2.3	HBase command-line client.
hbase-rest-server	1.2.3	Service providing a RESTful HTTP endpoint for HBase.

Component	Version	Description
hbase-thrift-server	1.2.3	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.1.1-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.1.1-amzn-0	Hive command line client.
hive-metastore-server	2.1.1-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	2.1.1-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.11.0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.3.0	Oozie command-line client.

Component	Version	Description
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.157.1	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.157.1	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.1.0	Spark command-line clients.
spark-history-server	2.1.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.1.0	In-memory execution engine for YARN.
spark-yarn-slave	2.1.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.

Component	Version	Description
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.6.2	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.9	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.9	ZooKeeper command line client.

5.3.2 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.3.2 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.

Classifications	Description
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.

Classifications	Description
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.

Classifications	Description
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.

Classifications	Description
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.

Classifications	Description
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.3.1

5.3.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.3.1	emr-5.3.0	emr-5.2.3	emr-5.2.2
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.4	1.1.4	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.1	2.1.1	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.0	2.1.0
Hudi	-	-	-	-

	emr-5.3.1	emr-5.3.0	emr-5.2.3	emr-5.2.2
Hue	3.11.0	3.11.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.157.1	0.157.1	0.157.1	0.157.1
Spark	2.1.0	2.1.0	2.0.2	2.0.2
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.9

5.3.1 release notes

The following release notes include information for the Amazon EMR 5.3.1 release. Changes are relative to the Amazon EMR 5.3.0 release.

Release date: February 7, 2017

Minor changes to backport Zeppelin patches and update the default AMI for Amazon EMR.

5.3.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.2.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.2.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.

Component	Version	Description
emrfs	2.14.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.1.4	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-1	HTTP endpoint for HDFS operations.

Component	Version	Description
hadoop-kms-server	2.7.3-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.3	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.3	Service for serving one or more HBase regions.
hbase-client	1.2.3	HBase command-line client.
hbase-rest-server	1.2.3	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.3	Service providing a Thrift endpoint to HBase.

Component	Version	Description
hcatalog-client	2.1.1-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.1.1-amzn-0	Hive command line client.
hive-metastore-server	2.1.1-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	2.1.1-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.11.0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.

Component	Version	Description
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.157.1	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.157.1	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.1.0	Spark command-line clients.
spark-history-server	2.1.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.1.0	In-memory execution engine for YARN.
spark-yarn-slave	2.1.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.

Component	Version	Description
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.6.2	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.9	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.9	ZooKeeper command line client.

5.3.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.3.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.

Classifications	Description
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.

Classifications	Description
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.

Classifications	Description
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.

Classifications	Description
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.

Classifications	Description
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.3.0

5.3.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.3.0	**emr-5.2.3**	**emr-5.2.2**	**emr-5.2.1**
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.4	1.1.3	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.1	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-

	emr-5.3.0	emr-5.2.3	emr-5.2.2	emr-5.2.1
Hue	3.11.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.157.1	0.157.1	0.157.1	0.157.1
Spark	2.1.0	2.0.2	2.0.2	2.0.2
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.9

5.3.0 release notes

The following release notes include information for the Amazon EMR 5.3.0 release. Changes are relative to the Amazon EMR 5.2.1 release.

Release date: January 26, 2017

Upgrades

- Upgraded to Hive 2.1.1
- Upgraded to Hue 3.11.0
- Upgraded to Spark 2.1.0
- Upgraded to Oozie 4.3.0
- Upgraded to Flink 1.1.4

Changes and enhancements

- Added a patch to Hue that allows you to use the `interpreters_shown_on_wheel` setting to configure what interpreters to show first on the Notebook selection wheel, regardless of their ordering in the `hue.ini` file.
- Added the `hive-parquet-logging` configuration classification, which you can use to configure values in the Hive `parquet-logging.properties` file.

5.3.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.2.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.2.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.14.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.1.4	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.

Component	Version	Description
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.3	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.

Component	Version	Description
hbase-region-server	1.2.3	Service for serving one or more HBase regions.
hbase-client	1.2.3	HBase command-line client.
hbase-rest-server	1.2.3	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.3	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.1.1-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.1.1-amzn-0	Hive command line client.
hive-metastore-server	2.1.1-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	2.1.1-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.11.0	Web application for analyzing data using Hadoop ecosystem applications

Component	Version	Description
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.52	MySQL database server.
oozie-client	4.3.0	Oozie command-line client.
oozie-server	4.3.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.157.1	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.157.1	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.1.0	Spark command-line clients.
spark-history-server	2.1.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.1.0	In-memory execution engine for YARN.

Component	Version	Description
spark-yarn-slave	2.1.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.6.2	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.9	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.9	ZooKeeper command line client.

5.3.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.3.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.

Classifications	Description
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.

Classifications	Description
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-parquet-logging	Change values in Hive's parquet-logging.pr operties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file

Classifications	Description
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.

Classifications	Description
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.

Classifications	Description
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.2.3

5.2.3 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.2.3	emr-5.2.2	emr-5.2.1	emr-5.2.0
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	1.1.3	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3

	emr-5.2.3	emr-5.2.2	emr-5.2.1	emr-5.2.0
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEn terpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.157.1	0.157.1	0.157.1	0.152.3
Spark	2.0.2	2.0.2	2.0.2	2.0.2
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.2.3	emr-5.2.2	emr-5.2.1	emr-5.2.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.8

5.2.3 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

> **Important**
>
> In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

5.2.3 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.2.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.2.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.13.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.1.3	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.

Component	Version	Description
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.3	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.

Component	Version	Description
hbase-region-server	1.2.3	Service for serving one or more HBase regions.
hbase-client	1.2.3	HBase command-line client.
hbase-rest-server	1.2.3	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.3	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.1.0-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.1.0-amzn-0	Hive command line client.
hive-metastore-server	2.1.0-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	2.1.0-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.10.0-amzn-0	Web application for analyzing data using Hadoop ecosystem applications

Component	Version	Description
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.52	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.157.1	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.157.1	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.0.2	Spark command-line clients.
spark-history-server	2.0.2	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.0.2	In-memory execution engine for YARN.

Component	Version	Description
spark-yarn-slave	2.0.2	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.23	Apache HTTP server.
zeppelin-server	0.6.2	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.9	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.9	ZooKeeper command line client.

5.2.3 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.2.3 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.

Classifications	Description
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.

Classifications	Description
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file

Classifications	Description
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.

Classifications	Description
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.

Classifications	Description
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.2.2

5.2.2 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.2.2	emr-5.2.1	emr-5.2.0	emr-5.1.1
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	1.1.3	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-

	emr-5.2.2	emr-5.2.1	emr-5.2.0	emr-5.1.1
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.157.1	0.157.1	0.152.3	0.152.3
Spark	2.0.2	2.0.2	2.0.2	2.0.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.8	3.4.8

5.2.2 release notes

The following release notes include information for the Amazon EMR 5.2.2 release. Changes are relative to the Amazon EMR 5.2.1 release.

Release date: May 2, 2017

Known issues resolved from previous releases

- Backported SPARK-194459, which addresses an issue where reading from an ORC table with char/varchar columns can fail.

5.2.2 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.2.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.2.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.

Component	Version	Description
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.13.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.1.3	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS service for tracking file names and block locations.

Component	Version	Description
hadoop-httpfs-server	2.7.3-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.3	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.3	Service for serving one or more HBase regions.
hbase-client	1.2.3	HBase command-line client.
hbase-rest-server	1.2.3	Service providing a RESTful HTTP endpoint for HBase.

Component	Version	Description
hbase-thrift-server	1.2.3	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.1.0-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.1.0-amzn-0	Hive command line client.
hive-metastore-server	2.1.0-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	2.1.0-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.10.0-amzn-0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.52	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.

Component	Version	Description
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.157.1	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.157.1	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.0.2	Spark command-line clients.
spark-history-server	2.0.2	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.0.2	In-memory execution engine for YARN.
spark-yarn-slave	2.0.2	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.

Component	Version	Description
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.23	Apache HTTP server.
zeppelin-server	0.6.2	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.9	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.9	ZooKeeper command line client.

5.2.2 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.2.2 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.

Classifications	Description
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.

Classifications	Description
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.

Classifications	Description
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.

Classifications	Description
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.

Classifications	Description
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.2.1

5.2.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.2.1	**emr-5.2.0**	**emr-5.1.1**	**emr-5.1.0**
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	1.1.3	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0

	emr-5.2.1	emr-5.2.0	emr-5.1.1	emr-5.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.157.1	0.152.3	0.152.3	0.152.3
Spark	2.0.2	2.0.2	2.0.1	2.0.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.8	3.4.8	3.4.8

5.2.1 release notes

The following release notes include information for the Amazon EMR 5.2.1 release. Changes are relative to the Amazon EMR 5.2.0 release.

Release date: December 29, 2016

Upgrades

- Upgraded to Presto 0.157.1. For more information, see Presto Release Notes in the Presto documentation.
- Upgraded to Zookeeper 3.4.9. For more information, see ZooKeeper Release Notes in the Apache ZooKeeper documentation.

Changes and enhancements

- Added support for the Amazon EC2 m4.16xlarge instance type in Amazon EMR version 4.8.3 and later, excluding 5.0.0, 5.0.3, and 5.2.0.
- Amazon EMR releases are now based on Amazon Linux 2016.09. For more information, see https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/.
- The location of Flink and YARN configuration paths are now set by default in `/etc/default/flink` that you do not need to set the environment variables `FLINK_CONF_DIR` and `HADOOP_CONF_DIR` when running the `flink` or `yarn-session.sh` driver scripts to launch Flink jobs.
- Added support for FlinkKinesisConsumer class.

Known issues resolved from previous releases

- Fixed an issue in Hadoop where the ReplicationMonitor thread could get stuck for a long time because of a race between replication and deletion of the same file in a large cluster.
- Fixed an issue where ControlledJob#toString failed with a null pointer exception (NPE) when job status was not successfully updated.

5.2.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system

processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.2.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.2.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.13.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.1.3	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.

Component	Version	Description
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN service for managing containers on an individual node.

Component	Version	Description
hadoop-yarn-resour cemanager	2.7.3-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.3	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.3	Service for serving one or more HBase regions.
hbase-client	1.2.3	HBase command-line client.
hbase-rest-server	1.2.3	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.3	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.1.0-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP endpoint providing a REST interface to HCatalog.

Component	Version	Description
hive-client	2.1.0-amzn-0	Hive command line client.
hive-metastore-server	2.1.0-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	2.1.0-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.10.0-amzn-0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.52	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.157.1	Service for accepting queries and managing query execution among presto-wo rkers.

Component	Version	Description
presto-worker	0.157.1	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.0.2	Spark command-line clients.
spark-history-server	2.0.2	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.0.2	In-memory execution engine for YARN.
spark-yarn-slave	2.0.2	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.23	Apache HTTP server.
zeppelin-server	0.6.2	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.9	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.

Component	Version	Description
zookeeper-client	3.4.9	ZooKeeper command line client.

5.2.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see [Configure applications](#).

emr-5.2.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration

Classifications	Description
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-env	Change values in the Hive environment.

Classifications	Description
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.

Classifications	Description
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.

Classifications	Description
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.

Classifications	Description
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.2.0

5.2.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.2.0	emr-5.1.1	emr-5.1.0	emr-5.0.3
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75

	emr-5.2.0	emr-5.1.1	emr-5.1.0	emr-5.0.3
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	1.1.3	1.1.3	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0

	emr-5.2.0	emr-5.1.1	emr-5.1.0	emr-5.0.3
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.152.3	0.152.3	0.152.3	0.152.3
Spark	2.0.2	2.0.1	2.0.1	2.0.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

5.2.0 release notes

The following release notes include information for the Amazon EMR 5.2.0 release. Changes are relative to the Amazon EMR 5.1.0 release.

Release date: November 21, 2016

Changes and enhancements

- Added Amazon S3 storage mode for HBase.
- Enables you to specify an Amazon S3 location for the HBase rootdir. For more information, see HBase on Amazon S3.

Upgrades

- Upgraded to Spark 2.0.2

Known issues resolved from previous releases

- Fixed an issue with /mnt being constrained to 2 TB on EBS-only instance types.
- Fixed an issue with instance-controller and logpusher logs being output to their corresponding .out files instead of to their normal log4j-configured .log files, which rotate hourly. The .out files do not rotate, so this would eventually fill up the /emr partition. This issue only affects hardware virtual machine (HVM) instance types.

5.2.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.1.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.1.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.

Component	Version	Description
emrfs	2.12.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.1.3	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-0	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-0	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-0	HTTP endpoint for HDFS operations.

Component	Version	Description
hadoop-kms-server	2.7.3-amzn-0	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-0	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-0	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.3	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.3	Service for serving one or more HBase regions.
hbase-client	1.2.3	HBase command-line client.
hbase-rest-server	1.2.3	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.3	Service providing a Thrift endpoint to HBase.

Component	Version	Description
hcatalog-client	2.1.0-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.1.0-amzn-0	Hive command line client.
hive-metastore-server	2.1.0-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	2.1.0-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.10.0-amzn-0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.52	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.

Component	Version	Description
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.152.3	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.152.3	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.0.2	Spark command-line clients.
spark-history-server	2.0.2	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.0.2	In-memory execution engine for YARN.
spark-yarn-slave	2.0.2	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.

Component	Version	Description
webserver	2.4.23	Apache HTTP server.
zeppelin-server	0.6.2	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.8	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.8	ZooKeeper command line client.

5.2.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.2.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.

Classifications	Description
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase	Amazon EMR-curated settings for Apache HBase.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.

Classifications	Description
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.

Classifications	Description
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.

Classifications	Description
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.

Classifications	Description
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.1.1

5.1.1 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.2
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	1.1.3	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.2
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.152.3	0.152.3	0.152.3	0.150
Spark	2.0.1	2.0.1	2.0.1	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

5.1.1 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

> **Important**
>
> In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

5.1.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.1.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.1.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.

Component	Version	Description
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.11.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.1.3	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-0	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-0	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFS service for tracking file names and block locations.

Component	Version	Description
hadoop-httpfs-server	2.7.3-amzn-0	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-0	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-0	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-0	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.3	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.3	Service for serving one or more HBase regions.
hbase-client	1.2.3	HBase command-line client.
hbase-rest-server	1.2.3	Service providing a RESTful HTTP endpoint for HBase.

Component	Version	Description
hbase-thrift-server	1.2.3	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.1.0-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.1.0-amzn-0	Hive command line client.
hive-metastore-server	2.1.0-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	2.1.0-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.10.0-amzn-0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.52	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.

Component	Version	Description
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.152.3	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.152.3	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.0.1	Spark command-line clients.
spark-history-server	2.0.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.0.1	In-memory execution engine for YARN.
spark-yarn-slave	2.0.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.

Component	Version	Description
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.23	Apache HTTP server.
zeppelin-server	0.6.2	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.8	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.8	ZooKeeper command line client.

5.1.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.1.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.

Classifications	Description
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.

Classifications	Description
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.

Classifications	Description
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.

Classifications	Description
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file

Classifications	Description
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.1.0

5.1.0 application versions

The following applications are supported in this release: Flink, Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases

- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.1.0	emr-5.0.3	emr-5.0.2	emr-5.0.1
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.2	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.2	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-

	emr-5.1.0	emr-5.0.3	emr-5.0.2	emr-5.0.1
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.152.3	0.152.3	0.150	0.150
Spark	2.0.1	2.0.1	2.0.0	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.1	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

5.1.0 release notes

The following release notes include information for the Amazon EMR 5.1.0 release. Changes are relative to the Amazon EMR 5.0.3 release.

Release date: November 03, 2016

Changes and enhancements

- Added support for Flink 1.1.3.
- Presto has been added as an option in the notebook section of Hue.

Upgrades

- Upgraded to HBase 1.2.3
- Upgraded to Zeppelin 0.6.2

Known issues resolved from previous releases

- Fixed an issue with Tez queries on Amazon S3 with ORC files did not perform as well as earlier Amazon EMR 4.x versions.

5.1.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.1.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.1.0	Extra convenience libraries for the Hadoop ecosystem.

Component	Version	Description
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.11.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
flink-client	1.1.3	Apache Flink command line client scripts and applications.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-0	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-0	HDFS command-line client and library

Component	Version	Description
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-0	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-0	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-0	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-0	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.3	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.3	Service for serving one or more HBase regions.
hbase-client	1.2.3	HBase command-line client.

Component	Version	Description
hbase-rest-server	1.2.3	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.3	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.1.0-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.1.0-amzn-0	Hive command line client.
hive-metastore-server	2.1.0-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	2.1.0-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.10.0-amzn-0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.52	MySQL database server.

Component	Version	Description
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.152.3	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.152.3	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.0.1	Spark command-line clients.
spark-history-server	2.0.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.0.1	In-memory execution engine for YARN.
spark-yarn-slave	2.0.1	Apache Spark libraries needed by YARN slaves.

Component	Version	Description
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.23	Apache HTTP server.
zeppelin-server	0.6.2	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.8	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.8	ZooKeeper command line client.

5.1.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.1.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.

Classifications	Description
emrfs-site	Change EMRFS settings.
flink-conf	Change flink-conf.yaml settings.
flink-log4j	Change Flink log4j.properties settings.
flink-log4j-yarn-session	Change Flink log4j-yarn-session.properties settings.
flink-log4j-cli	Change Flink log4j-cli.properties settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.

Classifications	Description
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.

Classifications	Description
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.

Classifications	Description
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.

Classifications	Description
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.0.3

5.0.3 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
Iceberg	-	-	-	-
JupyterEn terpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.152.3	0.150	0.150	0.150
Spark	2.0.1	2.0.0	2.0.0	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQ L)	-	-	-	-
Zeppelin	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

5.0.3 release notes

The following release notes include information for the Amazon EMR 5.0.3 release. Changes are relative to the Amazon EMR 5.0.0 release.

Release date: October 24, 2016

Upgrades

- Upgraded to Hadoop 2.7.3
- Upgraded to Presto 0.152.3, which includes support for the Presto web interface. You can access the Presto web interface on the Presto coordinator using port 8889. For more information about the Presto web interface, see Web Interface in the Presto documentation.
- Upgraded to Spark 2.0.1
- Amazon EMR releases are now based on Amazon Linux 2016.09. For more information, see https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/.

5.0.3 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.1.0	Amazon DynamoDB connector for Hadoop ecosystem applications.

Component	Version	Description
emr-goodies	2.1.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.10.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-0	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-0	HDFS command-line client and library

Component	Version	Description
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-0	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-0	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-0	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-0	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.2	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.2	Service for serving one or more HBase regions.
hbase-client	1.2.2	HBase command-line client.

Component	Version	Description
hbase-rest-server	1.2.2	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.2	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.1.0-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.1.0-amzn-0	Hive command line client.
hive-metastore-server	2.1.0-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	2.1.0-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.10.0-amzn-0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.52	MySQL database server.

Component	Version	Description
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.152.3	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.152.3	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.0.1	Spark command-line clients.
spark-history-server	2.0.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.0.1	In-memory execution engine for YARN.
spark-yarn-slave	2.0.1	Apache Spark libraries needed by YARN slaves.

Component	Version	Description
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.23	Apache HTTP server.
zeppelin-server	0.6.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.8	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.8	ZooKeeper command line client.

5.0.3 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.0.3 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.

Classifications	Description
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.

Classifications	Description
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.

Classifications	Description
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.

Classifications	Description
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.

Classifications	Description
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.0.2

5.0.2 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases

- [Application versions in Amazon EMR 5.x releases]
- [Application versions in Amazon EMR 4.x releases]

Application version information

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.152.3	0.150	0.150	0.150
Spark	2.0.1	2.0.0	2.0.0	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

5.0.2 release notes

5.0.2 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
emr-ddb	4.0.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.1.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.9.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.

Component	Version	Description
hadoop-client	2.7.2-amzn-3	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.2-amzn-3	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.2-amzn-3	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.2-amzn-3	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.2-amzn-3	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARN service for managing containers on an individual node.
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARN service for allocating and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Service for retrieving current and historical information for YARN applications.

Component	Version	Description
hbase-hmaster	1.2.2	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.2	Service for serving one or more HBase regions.
hbase-client	1.2.2	HBase command-line client.
hbase-rest-server	1.2.2	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.2	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.1.0-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.1.0-amzn-0	Hive command line client.
hive-metastore-server	2.1.0-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.

Component	Version	Description
hive-server	2.1.0-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.10.0-amzn-0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.46	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.150	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.150	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.0.0	Spark command-line clients.

Component	Version	Description
spark-history-server	2.0.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.0.0	In-memory execution engine for YARN.
spark-yarn-slave	2.0.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.23	Apache HTTP server.
zeppelin-server	0.6.1-SNAPSHOT	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.8	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.8	ZooKeeper command line client.

5.0.2 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.0.2 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.

Classifications	Description
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.

Classifications	Description
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.

Classifications	Description
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.

Classifications	Description
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.0.1

5.0.1 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.152.3	0.150	0.150	0.150
Spark	2.0.1	2.0.0	2.0.0	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

5.0.1 release notes

5.0.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.0.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.1.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.9.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati

Component	Version	Description
		ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.2-amzn-3	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.2-amzn-3	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.2-amzn-3	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.2-amzn-3	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.2-amzn-3	MapReduce execution engine libraries for running a MapReduce application.

Component	Version	Description
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.2-amzn-3	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.2	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.2	Service for serving one or more HBase regions.
hbase-client	1.2.2	HBase command-line client.
hbase-rest-server	1.2.2	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.2	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.1.0-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.

Component	Version	Description
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.1.0-amzn-0	Hive command line client.
hive-metastore-server	2.1.0-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	2.1.0-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.10.0-amzn-0	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.46	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API

Component	Version	Description
presto-coordinator	0.150	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.150	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.0.0	Spark command-line clients.
spark-history-server	2.0.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.0.0	In-memory execution engine for YARN.
spark-yarn-slave	2.0.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.23	Apache HTTP server.
zeppelin-server	0.6.1-SNAPSHOT	Web-based notebook that enables interactive data analytics.

Component	Version	Description
zookeeper-server	3.4.8	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.8	ZooKeeper command line client.

5.0.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.0.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration

Classifications	Description
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.

Classifications	Description
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.

Classifications	Description
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.

Classifications	Description
presto-connector-postgresql	Change values in Presto's postgresql.properties file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.

Classifications	Description
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 5.0.0

5.0.0 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie, Phoenix, Pig, Presto, Spark, Sqoop, Tez, Zeppelin, and ZooKeeper.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEn terpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.152.3	0.150	0.150	0.150
Spark	2.0.1	2.0.0	2.0.0	2.0.0

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

5.0.0 release notes

Release date: July 27, 2016

Upgrades

- Upgraded to Hive 2.1
- Upgraded to Presto 0.150
- Upgraded to Spark 2.0
- Upgraded to Hue 3.10.0
- Upgraded to Pig 0.16.0
- Upgraded to Tez 0.8.4
- Upgraded to Zeppelin 0.6.1

Changes and enhancements

- Amazon EMR supports the latest open-source versions of Hive (version 2.1) and Pig (version 0.16.0). If you have used Hive or Pig on Amazon EMR in the past, this may affect some use cases. For more information, see Hive and Pig.
- The default execution engine for Hive and Pig is now Tez. To change this, you would edit the appropriate values in the `hive-site` and `pig-properties` configuration classifications, respectively.

- An enhanced step debugging feature was added, which allows you to see the root cause of step failures if the service can determine the cause. For more information, see Enhanced Step Debugging in the Amazon EMR Management Guide.
- Applications that previously ended with "-Sandbox" no longer have that suffix. This may break your automation, for example, if you are using scripts to launch clusters with these applications. The following table shows application names in Amazon EMR 4.7.2 versus Amazon EMR 5.0.0.

Application name changes

Amazon EMR 4.7.2	Amazon EMR 5.0.0
Oozie-Sandbox	Oozie
Presto-Sandbox	Presto
Sqoop-Sandbox	Sqoop
Zeppelin-Sandbox	Zeppelin
ZooKeeper-Sandbox	ZooKeeper

- Spark is now compiled for Scala 2.11.
- Java 8 is now the default JVM. All applications run using the Java 8 runtime. There are no changes to any applicationâ€™s byte code target. Most applications continue to target Java 7.
- Zeppelin now includes authentication features. For more information, see Zeppelin.
- Added support for security configurations, which allow you to create and apply encryption options more easily. For more information, see Data Encryption.

5.0.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2

has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.0.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.1.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.9.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.2-amzn-3	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.

Component	Version	Description
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.2-amzn-3	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.2-amzn-3	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.2-amzn-3	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.2-amzn-3	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.2-amzn-3	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.2	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.

Component	Version	Description
hbase-region-server	1.2.2	Service for serving one or more HBase regions.
hbase-client	1.2.2	HBase command-line client.
hbase-rest-server	1.2.2	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.2	Service providing a Thrift endpoint to HBase.
hcatalog-client	2.1.0-amzn-0	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP endpoint providing a REST interface to HCatalog.
hive-client	2.1.0-amzn-0	Hive command line client.
hive-metastore-server	2.1.0-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	2.1.0-amzn-0	Service for accepting Hive queries as web requests.
hue-server	3.10.0-amzn-0	Web application for analyzing data using Hadoop ecosystem applications

Component	Version	Description
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.46	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.150	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.150	Service for executing pieces of a query.
pig-client	0.16.0-amzn-0	Pig command-line client.
spark-client	2.0.0	Spark command-line clients.
spark-history-server	2.0.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	2.0.0	In-memory execution engine for YARN.

Component	Version	Description
spark-yarn-slave	2.0.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.23	Apache HTTP server.
zeppelin-server	0.6.1-SNAPSHOT	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.8	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.8	ZooKeeper command line client.

5.0.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-5.0.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.

Classifications	Description
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.

Classifications	Description
hcatalog-webhcat-log4j2	Change values in HCatalog WebHCat's log4j2.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-beeline-log4j2	Change values in Hive's beeline-log4j2.pro perties file.
hive-env	Change values in the Hive environment.
hive-exec-log4j2	Change values in Hive's hive-exec-log4j2.p roperties file.
hive-llap-daemon-log4j2	Change values in Hive's llap-daemon-log4j2 .properties file.
hive-log4j2	Change values in Hive's hive-log4j2.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.

Classifications	Description
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.

Classifications	Description
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-hive-site	Change values in Spark's hive-site.xml file
spark-log4j	Change values in Spark's log4j.properties file.

Classifications	Description
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR 4.x release versions

This section contains application versions, release notes, component versions, and configuration classifications available in each Amazon EMR 4.x release version.

When you launch a cluster, you can choose from multiple releases of Amazon EMR. This allows you to test and use application versions that fit your compatibility requirements. You specify the release number with the *release label*. Release labels are in the form `emr-x.x.x`. For example, `emr-7.2.0`.

New Amazon EMR releases are made available in different Regions over a period of several days, beginning with the first Region on the initial release date. The latest release version may not be available in your Region during this period.

For a comprehensive table of application versions in every Amazon EMR 4.x release, see Application versions in Amazon EMR 4.x releases.

Topics

Application versions in Amazon EMR 4.x releases

For a comprehensive table that lists the application versions available in each Amazon EMR 4.x release, open Application versions in Amazon EMR 4.x releases in your browser.

Differences in Amazon EMR 4.x release versions

Documentation for Amazon EMR features in the *Amazon EMR Management Guide* specify the Amazon EMR release version that a feature became available, as well as applicable differences between Amazon EMR features dating back to 4.0.0.

Beginning with Amazon EMR release version 5.0.0, some applications got a significant version upgrade that changed installation or operational details, and others were promoted from sandbox applications to native applications. Each topic in this section provides notable application-specific differences when using Amazon EMR 4.x release versions.

Topics

Sandbox applications

When using Amazon EMR 4.x release versions, some applications are considered *sandbox* applications. Sandbox applications are early versions of the application that we made available at the time of the initial Amazon EMR release because of demand. You can use the console, AWS CLI, or API to have Amazon EMR install sandbox applications in the same way as native applications, but sandbox applications have limited support and documentation. Sandbox applications became native, fully supported applications in Amazon EMR release version 5.0.0 and later. The following are sandbox applications in Amazon EMR 4.x release versions:

- Oozie
- Presto
- Sqoop
- Zeppelin
- ZooKeeper

When you install sandbox applications, the application names are denoted with the suffix `-sandbox`. For example, to install the sandbox version of *Presto*, use `Presto-sandbox`. Installation may take longer than it does for a fully supported application. The version numbers listed for each application in this section correspond to the community version of the application.

Oozie (sandbox versions)

Oozie is available as a sandbox application beginning with Amazon EMR release version 4.1.0.

Oozie examples are not installed by default using the sandbox versions. To install the examples, SSH to the master node and run `install-oozie-examples`.

Oozie-Sandbox version information

Amazon EMR Release label	Oozie-Sandbox Version	Components installed with Oozie-Sandbox
emr-4.9.6	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.9.5	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resour

Amazon EMR Release label	Oozie-Sandbox Version	Components installed with Oozie-Sandbox
		cemanager, oozie-client, oozie-server
emr-4.9.4	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.9.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Amazon EMR Release label	Oozie-Sandbox Version	Components installed with Oozie-Sandbox
emr-4.9.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.9.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Amazon EMR Release label	Oozie-Sandbox Version	Components installed with Oozie-Sandbox
emr-4.8.5	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.8.4	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Amazon EMR Release label	Oozie-Sandbox Version	Components installed with Oozie-Sandbox
emr-4.8.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.8.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Amazon EMR Release label	Oozie-Sandbox Version	Components installed with Oozie-Sandbox
emr-4.8.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.8.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Amazon EMR Release label	Oozie-Sandbox Version	Components installed with Oozie-Sandbox
emr-4.7.4	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.7.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Amazon EMR Release label	Oozie-Sandbox Version	Components installed with Oozie-Sandbox
emr-4.7.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.7.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Amazon EMR Release label	Oozie-Sandbox Version	Components installed with Oozie-Sandbox
emr-4.7.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.6.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.6.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Amazon EMR Release label	Oozie-Sandbox Version	Components installed with Oozie-Sandbox
emr-4.5.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.4.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.3.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Amazon EMR Release label	Oozie-Sandbox Version	Components installed with Oozie-Sandbox
emr-4.2.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.1.0	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Presto (sandbox versions)

Presto is available as a sandbox application beginning with Amazon EMR release version 4.1.0.

Presto-Sandbox version information

Amazon EMR Release label	Presto-Sandbox Version	Components installed with Presto-Sandbox
emr-4.9.6	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode,

Amazon EMR Release label	Presto-Sandbox Version	Components installed with Presto-Sandbox
		hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.9.5	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.9.4	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto-Sandbox Version	Components installed with Presto-Sandbox
emr-4.9.3	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.9.2	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.9.1	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto-Sandbox Version	Components installed with Presto-Sandbox
emr-4.8.5	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.8.4	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.8.3	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto-Sandbox Version	Components installed with Presto-Sandbox
emr-4.8.2	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.8.1	0.151	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.8.0	0.151	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto-Sandbox Version	Components installed with Presto-Sandbox
emr-4.7.4	0.148	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.7.3	0.148	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.7.2	0.148	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto-Sandbox Version	Components installed with Presto-Sandbox
emr-4.7.1	0.147	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.7.0	0.147	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.6.1	0.143	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto-Sandbox Version	Components installed with Presto-Sandbox
emr-4.6.0	0.143	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker
emr-4.5.0	0.140	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker
emr-4.4.0	0.136	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto-Sandbox Version	Components installed with Presto-Sandbox
emr-4.3.0	0.130	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker
emr-4.2.0	0.125	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker
emr-4.1.0	0.119	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker

Sqoop (sandbox versions)

Sqoop is available as a sandbox application beginning with Amazon EMR release version 4.4.0.

Sqoop-Sandbox version information

Amazon EMR Release label	Sqoop-Sandbox Version	Components installed with Sqoop-Sandbox
emr-4.9.6	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.9.5	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.9.4	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resour

Amazon EMR Release label	Sqoop-Sandbox Version	Components installed with Sqoop-Sandbox
		cemanager, mysql-server, sqoop-client
emr-4.9.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.9.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop-Sandbox Version	Components installed with Sqoop-Sandbox
emr-4.9.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.8.5	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.8.4	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop-Sandbox Version	Components installed with Sqoop-Sandbox
emr-4.8.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.8.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.8.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop-Sandbox Version	Components installed with Sqoop-Sandbox
emr-4.8.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.7.4	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.7.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop-Sandbox Version	Components installed with Sqoop-Sandbox
emr-4.7.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.7.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.7.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop-Sandbox Version	Components installed with Sqoop-Sandbox
emr-4.6.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.6.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.5.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, sqoop-client

Amazon EMR Release label	Sqoop-Sandbox Version	Components installed with Sqoop-Sandbox
emr-4.4.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, sqoop-client

Zeppelin (sandbox versions)

Zeppelin is available as a sandbox application beginning with Amazon EMR release version 4.1.0.

Zeppelin-Sandbox version information

Amazon EMR Release label	Zeppelin-Sandbox Version	Components installed with Zeppelin-Sandbox
emr-4.9.6	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.9.5	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-nameno

Amazon EMR Release label	Zeppelin-Sandbox Version	Components installed with Zeppelin-Sandbox
		de, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.9.4	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.9.3	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin-Sandbox Version	Components installed with Zeppelin-Sandbox
emr-4.9.2	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.9.1	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin-Sandbox Version	Components installed with Zeppelin-Sandbox
emr-4.8.5	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.8.4	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin-Sandbox Version	Components installed with Zeppelin-Sandbox
emr-4.8.3	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.8.2	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin-Sandbox Version	Components installed with Zeppelin-Sandbox
emr-4.8.1	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.8.0	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin-Sandbox Version	Components installed with Zeppelin-Sandbox
emr-4.7.4	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.7.3	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin-Sandbox Version	Components installed with Zeppelin-Sandbox
emr-4.7.2	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.7.1	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin-Sandbox Version	Components installed with Zeppelin-Sandbox
emr-4.7.0	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.6.1	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin-Sandbox Version	Components installed with Zeppelin-Sandbox
emr-4.6.0	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.5.0	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin-Sandbox Version	Components installed with Zeppelin-Sandbox
emr-4.4.0	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.3.0	0.5.5	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin-Sandbox Version	Components installed with Zeppelin-Sandbox
emr-4.2.0	0.5.5	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.1.0	0.6.0-SNAPSHOT	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

ZooKeeper (sandbox versions)

Zookeeper is available as a sandbox application beginning with Amazon EMR release version 4.6.0.

ZooKeeper-Sandbox version information

Amazon EMR Release label	ZooKeeper-Sandbox Version	Components installed with ZooKeeper-Sandbox
emr-4.9.6	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datano

Amazon EMR Release label	ZooKeeper-Sandbox Version	Components installed with ZooKeeper-Sandbox
		de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.9.5	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.9.4	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper-Sandbox Version	Components installed with ZooKeeper-Sandbox
emr-4.9.3	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.9.2	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.9.1	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper-Sandbox Version	Components installed with ZooKeeper-Sandbox
emr-4.8.5	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.8.4	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.8.3	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper-Sandbox Version	Components installed with ZooKeeper-Sandbox
emr-4.8.2	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.8.1	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.8.0	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper-Sandbox Version	Components installed with ZooKeeper-Sandbox
emr-4.7.4	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.7.3	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.7.2	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper-Sandbox Version	Components installed with ZooKeeper-Sandbox
emr-4.7.1	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.7.0	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.6.1	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper-Sandbox Version	Components installed with ZooKeeper-Sandbox
emr-4.6.0	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Considerations for using Hive on Amazon EMR 4.x

This section covers differences to consider when using Hive version 1.0.0 on Amazon EMR 4.x release versions, as compared to Hive 2.x on Amazon EMR 5.x release versions.

ACID transactions not supported

Hive on Amazon EMR 4.x release versions does not support ACID transactions with Hive data stored in Amazon S3 when using 4.x release versions. If you try to create a transactional table in Amazon S3, an exception occurs.

Reading and writing to tables in Amazon S3

Hive on Amazon EMR 4.x release versions can write directly to Amazon S3 without using temporary files. This improves performance, but a consequence is that you cannot read and write to the same table in Amazon S3 within the same Hive statement. A workaround is to create and use a temporary table in HDFS.

The following example shows how to use multiple Hive statements to update a table in Amazon S3. The statements create a temporary table in HDFS named `tmp` based on a table in Amazon S3 named `my_s3_table`. The table in Amazon S3 is then updated with the contents of the temporary table.

```
CREATE TEMPORARY TABLE tmp LIKE my_s3_table;
```

```
INSERT OVERWRITE TABLE tmp SELECT ....;
INSERT OVERWRITE TABLE my_s3_table SELECT * FROM tmp;
```

Log4j vs. Log4j 2

Hive on Amazon EMR 4.x release versions uses Log4j. Beginning with version 5.0.0, Log4j 2 is the default. These versions may require different logging configurations. See Apache Log4j 2 for details.

MapReduce is the default execution engine

Hive on Amazon EMR 4.x release versions uses MapReduce as the default execution engine. Beginning with Amazon EMR version 5.0.0, Tez is the default, which provides improved performance for most workflows.

Hive authorization

Hive on Amazon EMR 4.x release versions supports Hive authorization for HDFS but not for EMRFS and Amazon S3. Amazon EMR clusters run with authorization disabled by default.

Hive file merge behavior with Amazon S3

Hive on Amazon EMR 4.x release versions merges small files at the end of a map-only job if `hive.merge.mapfiles` is `true`. A merge is triggered only if the average output size of the job is less than the `hive.merge.smallfiles.avgsize` setting. Amazon EMR Hive has exactly the same behavior if the final output path is in HDFS. If the output path is in Amazon S3, however, the `hive.merge.smallfiles.avgsize` parameter is ignored. In that situation, the merge task is always triggered if `hive.merge.mapfiles` is set to `true`.

Considerations for using Pig on Amazon EMR 4.x

Pig version 0.14.0 is installed on clusters created using Amazon EMR 4.x release versions. Pig was upgraded to version 0.16.0 in Amazon EMR 5.0.0. Significant differences are covered below.

Different default execution engine

Pig version 0.14.0 on Amazon EMR 4.x release versions uses MapReduce as the default execution engine. Pig 0.16.0 and later use Apache Tez. You can explicitly set `exectype=mapreduce` in the `pig-properties` configuration classification to use MapReduce.

Dropped Pig user-defined functions (UDFs)

Custom UDFs that were available in Pig on Amazon EMR 4.x release versions were dropped beginning with Pig 0.16.0. Most of the UDFs have equivalent functions you can use instead. The following table lists dropped UDFs and equivalent functions. For more information, see Built-in functions on the Apache Pig site.

Dropped UDF	Equivalent function
FORMAT_DT(dtformat, date)	GetHour(date), GetMinute(date), GetMonth(date), GetSecond(date), GetWeek(date), GetYear(date), GetDay(date)
EXTRACT(string, pattern)	REGEX_EXTRACT_ALL(string, pattern)
REPLACE(string, pattern, replacement)	REPLACE(string, pattern, replacement)
DATE_TIME()	ToDate()
DURATION(dt, dt2)	WeeksBetween(dt, dt2), YearsBetween(dt, dt2), SecondsBetween(dt, dt2), MonthsBetween(dt, dt2), MinutesBetween(dt, dt2), HoursBetween(dt, dt2)
EXTRACT_DT(format, date)	GetHour(date), GetMinute(date), GetMonth(date), GetSecond(date), GetWeek(date), GetYear(date), GetDay(date)
OFFSET_DT(date, duration)	AddDuration(date, duration), SubtractDuration(date, duration)
PERIOD(dt, dt2)	WeeksBetween(dt, dt2), YearsBetween(dt, dt2), SecondsBetween(dt, dt2), MonthsBetween(dt, dt2), MinutesBetween(dt, dt2), HoursBetween(dt, dt2)
CAPITALIZE(string)	UCFIRST(string)
CONCAT_WITH()	CONCAT()

Dropped UDF	Equivalent function
INDEX_OF()	INDEXOF()
LAST_INDEX_OF()	LAST_INDEXOF()
SPLIT_ON_REGEX()	STRSPLT()
UNCAPITALIZE()	LCFIRST()

The following UDFs were dropped with no equivalent: FORMAT(), LOCAL_DATE(), LOCAL_TIME(), CENTER(), LEFT_PAD(), REPEAT(), REPLACE_ONCE(), RIGHT_PAD(), STRIP(), STRIP_END(), STRIP_START(), SWAP_CASE().

Discontinued Grunt commands

Some Grunt commands were discontinued beginning with Pig 0.16.0. The following table lists Grunt commands in Pig 0.14.0 and the equivalent commands in the current version, where applicable.

Pig 0.14.0 and equivalent current Grunt commands

Pig 0.14.0 Grunt command	Pig Grunt command in 0.16.0 and later
cat <non-hdfs-path>)	fs -cat <non-hdfs-path>;
cd <non-hdfs-path>;	No equivalent
ls <non-hdfs-path>;	fs -ls <non-hdfs-path>;
move <non-hdfs-path> <non-hdfs-path>;	fs -mv <non-hdfs-path> <non-hdfs-path>;
copy <non-hdfs-path> <non-hdfs-path>;	fs -cp <non-hdfs-path> <non-hdfs-path>;
copyToLocal <non-hdfs-path> <local-path>;	fs -copyToLocal <non-hdfs-path> <local-path>;
copyFromLocal <local-path> <non-hdfs-path>;	fs -copyFromLocal <local-path> <non-hdfs-path>;
mkdir <non-hdfs-path>;	fs -mkdir <non-hdfs-path>;

Pig 0.14.0 Grunt command	Pig Grunt command in 0.16.0 and later
rm <non-hdfs-path>;	fs -rm -r -skipTrash <non-hdfs-path>;
rmf <non-hdfs-path>;	fs -rm -r -skipTrash <non-hdfs-path>;

Capability removed for non-HDFS home directories

Pig 0.14.0 on Amazon EMR 4.x release versions has two mechanisms to allow users other than the `hadoop` user, who don't have home directories, to run Pig scripts. The first mechanism is an automatic fallback that sets the initial working directory to the root directory if the home directory doesn't exist. The second is a `pig.initial.fs.name` property that allows you to change the initial working directory.

These mechanisms are not available beginning with Amazon EMR version 5.0.0, and users must have a home directory on HDFS. This doesn't apply to the `hadoop` user because a home directory is provisioned at launch. Scripts run using Hadoop jar steps default to the Hadoop user unless another user is explicitly specified using `command-runner.jar`.

Amazon EMR release 4.9.6

4.9.6 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Phoenix, Pig, Presto-Sandbox, Spark, Sqoop-Sandbox, Tez, Zeppelin-Sandbox, and ZooKeeper-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.9.6	emr-4.9.5	emr-4.9.4	emr-4.9.3
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-4.9.6	emr-4.9.5	emr-4.9.4	emr-4.9.3
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

4.9.6 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

> **Important**
>
> In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

4.9.6 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.3.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.2.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.3.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.

Component	Version	Description
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.17.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-2	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP endpoint for HDFS operations.

Component	Version	Description
hadoop-kms-server	2.7.3-amzn-2	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-2	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-2	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.2	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.2	Service for serving one or more HBase regions.
hbase-client	1.2.2	HBase command-line client.
hbase-rest-server	1.2.2	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.2	Service providing a Thrift endpoint to HBase.

Component	Version	Description
hcatalog-client	1.0.0-amzn-9	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	1.0.0-amzn-9	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP endpoint providing a REST interface to HCatalog.
hive-client	1.0.0-amzn-9	Hive command line client.
hive-metastore-server	1.0.0-amzn-9	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	1.0.0-amzn-9	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-7	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.

Component	Version	Description
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.157.1	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.157.1	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.6.3	Spark command-line clients.
spark-history-server	1.6.3	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.6.3	In-memory execution engine for YARN.
spark-yarn-slave	1.6.3	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.

Component	Version	Description
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.6.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.9	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.9	ZooKeeper command line client.

4.9.6 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-4.9.6 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.

Classifications	Description
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j	Change values in HCatalog WebHCat's log4j.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.

Classifications	Description
hive-env	Change values in the Hive environment.
hive-exec-log4j	Change values in Hive's hive-exec-log4j.pr operties file.
hive-log4j	Change values in Hive's hive-log4j.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.

Classifications	Description
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.

Classifications	Description
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.

Classifications	Description
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 4.9.5

4.9.5 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Phoenix, Pig, Presto-Sandbox, Spark, Sqoop-Sandbox, Tez, Zeppelin-Sandbox, and ZooKeeper-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.9.5	emr-4.9.4	emr-4.9.3	emr-4.9.2
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked

	emr-4.9.5	emr-4.9.4	emr-4.9.3	emr-4.9.2
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0

	emr-4.9.5	emr-4.9.4	emr-4.9.3	emr-4.9.2
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

4.9.5 release notes

The following release notes include information for Amazon EMR release 4.9.5. Changes are relative to 4.9.4.

Initial release date: August 29, 2018

Changes, enhancements, and resolved issues

- HBase

- This release addresses a potential security vulnerability.

4.9.5 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.3.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.2.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.3.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.17.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati

Component	Version	Description
		ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-2	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-2	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-2	MapReduce execution engine libraries for running a MapReduce application.

Component	Version	Description
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-2	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.2	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.2	Service for serving one or more HBase regions.
hbase-client	1.2.2	HBase command-line client.
hbase-rest-server	1.2.2	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.2	Service providing a Thrift endpoint to HBase.
hcatalog-client	1.0.0-amzn-9	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	1.0.0-amzn-9	Service providing HCatalog, a table and storage management layer for distributed applications.

Component	Version	Description
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP endpoint providing a REST interface to HCatalog.
hive-client	1.0.0-amzn-9	Hive command line client.
hive-metastore-server	1.0.0-amzn-9	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	1.0.0-amzn-9	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-7	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API

Component	Version	Description
presto-coordinator	0.157.1	Service for accepting queries and managing query execution among presto-workers.
presto-worker	0.157.1	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.6.3	Spark command-line clients.
spark-history-server	1.6.3	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.6.3	In-memory execution engine for YARN.
spark-yarn-slave	1.6.3	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.6.1	Web-based notebook that enables interactive data analytics.

Component	Version	Description
zookeeper-server	3.4.9	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.9	ZooKeeper command line client.

4.9.5 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-4.9.5 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration

Classifications	Description
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j	Change values in HCatalog WebHCat's log4j.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-env	Change values in the Hive environment.
hive-exec-log4j	Change values in Hive's hive-exec-log4j.pr operties file.
hive-log4j	Change values in Hive's hive-log4j.properties file.

Classifications	Description
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environment.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.

Classifications	Description
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.

Classifications	Description
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 4.9.4

4.9.4 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Phoenix, Pig, Presto-Sandbox, Spark, Sqoop-Sandbox, Tez, Zeppelin-Sandbox, and ZooKeeper-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.9.4	emr-4.9.3	emr-4.9.2	emr-4.9.1
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2

	emr-4.9.4	emr-4.9.3	emr-4.9.2	emr-4.9.1
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6

	emr-4.9.4	emr-4.9.3	emr-4.9.2	emr-4.9.1
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

4.9.4 release notes

The following release notes include information for Amazon EMR release 4.9.4. Changes are relative to 4.9.3.

Initial release date: March 29, 2018

Changes, enhancements, and resolved issues

- Updated the Amazon Linux kernel of the defaultAmazon Linux AMI for Amazon EMR to address potential vulnerabilities.

4.9.4 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
emr-ddb	4.3.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.2.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.3.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.17.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.

Component	Version	Description
hadoop-client	2.7.3-amzn-2	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-2	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-2	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-2	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service for retrieving current and historical information for YARN applications.

Component	Version	Description
hbase-hmaster	1.2.2	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.2	Service for serving one or more HBase regions.
hbase-client	1.2.2	HBase command-line client.
hbase-rest-server	1.2.2	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.2	Service providing a Thrift endpoint to HBase.
hcatalog-client	1.0.0-amzn-9	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	1.0.0-amzn-9	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP endpoint providing a REST interface to HCatalog.
hive-client	1.0.0-amzn-9	Hive command line client.
hive-metastore-server	1.0.0-amzn-9	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.

Component	Version	Description
hive-server	1.0.0-amzn-9	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-7	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.157.1	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.157.1	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.6.3	Spark command-line clients.

Component	Version	Description
spark-history-server	1.6.3	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.6.3	In-memory execution engine for YARN.
spark-yarn-slave	1.6.3	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.6.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.9	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.9	ZooKeeper command line client.

4.9.4 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-4.9.4 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.

Classifications	Description
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j	Change values in HCatalog WebHCat's log4j.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-env	Change values in the Hive environment.
hive-exec-log4j	Change values in Hive's hive-exec-log4j.pr operties file.
hive-log4j	Change values in Hive's hive-log4j.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.

Classifications	Description
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.

Classifications	Description
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.

Classifications	Description
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 4.9.3

4.9.3 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Phoenix, Pig, Presto-Sandbox, Spark, Sqoop-Sandbox, Tez, Zeppelin-Sandbox, and ZooKeeper-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases

- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.9.3	emr-4.9.2	emr-4.9.1	emr-4.8.5
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-

	emr-4.9.3	emr-4.9.2	emr-4.9.1	emr-4.8.5
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-

	emr-4.9.3	emr-4.9.2	emr-4.9.1	emr-4.8.5
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

4.9.3 release notes

The following release notes include information for the Amazon EMR 4.9.3 release. Changes are relative to the Amazon EMR 4.9.2 release.

Initial release date: January 22, 2018

Changes, enhancements, and resolved issues

- Updated the Amazon Linux kernel of the defaultAmazon Linux AMI for Amazon EMR to address vulnerabilities associated with speculative execution (CVE-2017-5715, CVE-2017-5753, and CVE-2017-5754). For more information, see https://aws.amazon.com/security/security-bulletins/AWS-2018-013/.

4.9.3 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.3.0	Amazon DynamoDB connector for Hadoop ecosystem applications.

Component	Version	Description
emr-goodies	2.2.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.3.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.17.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-2	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS command-line client and library

Component	Version	Description
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-2	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-2	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-2	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.2	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.2	Service for serving one or more HBase regions.
hbase-client	1.2.2	HBase command-line client.

Component	Version	Description
hbase-rest-server	1.2.2	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.2	Service providing a Thrift endpoint to HBase.
hcatalog-client	1.0.0-amzn-9	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	1.0.0-amzn-9	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP endpoint providing a REST interface to HCatalog.
hive-client	1.0.0-amzn-9	Hive command line client.
hive-metastore-server	1.0.0-amzn-9	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	1.0.0-amzn-9	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-7	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.

Component	Version	Description
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.157.1	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.157.1	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.6.3	Spark command-line clients.
spark-history-server	1.6.3	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.6.3	In-memory execution engine for YARN.
spark-yarn-slave	1.6.3	Apache Spark libraries needed by YARN slaves.

Component	Version	Description
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.6.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.9	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.9	ZooKeeper command line client.

4.9.3 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-4.9.3 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.

Classifications	Description
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.

Classifications	Description
hcatalog-webhcat-log4j	Change values in HCatalog WebHCat's log4j.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-env	Change values in the Hive environment.
hive-exec-log4j	Change values in Hive's hive-exec-log4j.pr operties file.
hive-log4j	Change values in Hive's hive-log4j.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.

Classifications	Description
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.

Classifications	Description
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.

Classifications	Description
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 4.9.2

4.9.2 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Phoenix, Pig, Presto-Sandbox, Spark, Sqoop-Sandbox, Tez, Zeppelin-Sandbox, and ZooKeeper-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.9.2	emr-4.9.1	emr-4.8.5	emr-4.8.4
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-4.9.2	emr-4.9.1	emr-4.8.5	emr-4.8.4
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

4.9.2 release notes

The following release notes include information for the Amazon EMR 4.9.2 release. Changes are relative to the Amazon EMR 4.9.1 release.

Release date: July 13, 2017

Minor changes, bug fixes, and enhancements were made in this release.

4.9.2 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.3.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.2.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.3.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.

Component	Version	Description
emrfs	2.17.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-2	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-2	Cryptographic key management server based on Hadoop's KeyProvider API.

Component	Version	Description
hadoop-mapred	2.7.3-amzn-2	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-2	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.2	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.2	Service for serving one or more HBase regions.
hbase-client	1.2.2	HBase command-line client.
hbase-rest-server	1.2.2	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.2	Service providing a Thrift endpoint to HBase.
hcatalog-client	1.0.0-amzn-9	The 'hcat' command line client for manipulating hcatalog-server.

Component	Version	Description
hcatalog-server	1.0.0-amzn-9	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP endpoint providing a REST interface to HCatalog.
hive-client	1.0.0-amzn-9	Hive command line client.
hive-metastore-server	1.0.0-amzn-9	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	1.0.0-amzn-9	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-7	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client

Component	Version	Description
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.157.1	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.157.1	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.6.3	Spark command-line clients.
spark-history-server	1.6.3	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.6.3	In-memory execution engine for YARN.
spark-yarn-slave	1.6.3	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.

Component	Version	Description
zeppelin-server	0.6.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.9	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.9	ZooKeeper command line client.

4.9.2 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-4.9.2 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.

Classifications	Description
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j	Change values in HCatalog WebHCat's log4j.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-env	Change values in the Hive environment.

Classifications	Description
hive-exec-log4j	Change values in Hive's hive-exec-log4j.pr operties file.
hive-log4j	Change values in Hive's hive-log4j.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.

Classifications	Description
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.

Classifications	Description
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.

Classifications	Description
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 4.9.1

4.9.1 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Phoenix, Pig, Presto-Sandbox, Spark, Sqoop-Sandbox, Tez, Zeppelin-Sandbox, and ZooKeeper-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.9.1	emr-4.8.5	emr-4.8.4	emr-4.8.3
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-

	emr-4.9.1	emr-4.8.5	emr-4.8.4	emr-4.8.3
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-

	emr-4.9.1	emr-4.8.5	emr-4.8.4	emr-4.8.3
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

4.9.1 release notes

The following release notes include information for the Amazon EMR 4.9.1 release. Changes are relative to the Amazon EMR 4.8.4 release.

Release date: April 10, 2017

Known issues resolved from previous releases

- Backports of HIVE-9976 and HIVE-10106
- Fixed an issue in YARN where a large number of nodes (greater than 2,000) and containers (greater than 5,000) would cause an out-of-memory error, for example: `"Exception in thread main java.lang.OutOfMemoryError"`.

Changes and enhancements

- Amazon EMR releases are now based on Amazon Linux 2017.03. For more information, see https://aws.amazon.com/amazon-linux-ami/2017.03-release-notes/.
- Removed Python 2.6 from the Amazon EMR base Linux image. You can install Python 2.6 manually if necessary.

4.9.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.2.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.2.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.3.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.

Component	Version	Description
emrfs	2.15.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-2	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-2	Cryptographic key management server based on Hadoop's KeyProvider API.

Component	Version	Description
hadoop-mapred	2.7.3-amzn-2	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-2	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.2	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.2	Service for serving one or more HBase regions.
hbase-client	1.2.2	HBase command-line client.
hbase-rest-server	1.2.2	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.2	Service providing a Thrift endpoint to HBase.
hcatalog-client	1.0.0-amzn-9	The 'hcat' command line client for manipulating hcatalog-server.

Component	Version	Description
hcatalog-server	1.0.0-amzn-9	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP endpoint providing a REST interface to HCatalog.
hive-client	1.0.0-amzn-9	Hive command line client.
hive-metastore-server	1.0.0-amzn-9	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	1.0.0-amzn-9	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-7	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client

Component	Version	Description
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.157.1	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.157.1	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.6.3	Spark command-line clients.
spark-history-server	1.6.3	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.6.3	In-memory execution engine for YARN.
spark-yarn-slave	1.6.3	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.

Component	Version	Description
zeppelin-server	0.6.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.9	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.9	ZooKeeper command line client.

4.9.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-4.9.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.

Classifications	Description
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j	Change values in HCatalog WebHCat's log4j.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-env	Change values in the Hive environment.

Classifications	Description
hive-exec-log4j	Change values in Hive's hive-exec-log4j.pr operties file.
hive-log4j	Change values in Hive's hive-log4j.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.

Classifications	Description
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.

Classifications	Description
presto-connector-postgresql	Change values in Presto's postgresql.properties file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.

Classifications	Description
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 4.8.5

4.8.5 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Phoenix, Pig, Presto-Sandbox, Spark, Sqoop-Sandbox, Tez, Zeppelin-Sandbox, and ZooKeeper-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.8.5	emr-4.8.4	emr-4.8.3	emr-4.8.2
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-

	emr-4.8.5	emr-4.8.4	emr-4.8.3	emr-4.8.2
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-

	emr-4.8.5	emr-4.8.4	emr-4.8.3	emr-4.8.2
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.152.3
Spark	1.6.3	1.6.3	1.6.3	1.6.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.8

4.8.5 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

> **Important**
>
> In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

4.8.5 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.2.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.2.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.14.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.

Component	Version	Description
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN service for managing containers on an individual node.

Component	Version	Description
hadoop-yarn-resour cemanager	2.7.3-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.2	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.2	Service for serving one or more HBase regions.
hbase-client	1.2.2	HBase command-line client.
hbase-rest-server	1.2.2	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.2	Service providing a Thrift endpoint to HBase.
hcatalog-client	1.0.0-amzn-8	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	1.0.0-amzn-8	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	1.0.0-amzn-8	HTTP endpoint providing a REST interface to HCatalog.

Component	Version	Description
hive-client	1.0.0-amzn-8	Hive command line client.
hive-metastore-server	1.0.0-amzn-8	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	1.0.0-amzn-8	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-7	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.157.1	Service for accepting queries and managing query execution among presto-wo rkers.

Component	Version	Description
presto-worker	0.157.1	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.6.3	Spark command-line clients.
spark-history-server	1.6.3	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.6.3	In-memory execution engine for YARN.
spark-yarn-slave	1.6.3	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.6.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.9	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.

Component	Version	Description
zookeeper-client	3.4.9	ZooKeeper command line client.

4.8.5 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-4.8.5 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.

Classifications	Description
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j	Change values in HCatalog WebHCat's log4j.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-env	Change values in the Hive environment.
hive-exec-log4j	Change values in Hive's hive-exec-log4j.properties file.
hive-log4j	Change values in Hive's hive-log4j.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file

Classifications	Description
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.

Classifications	Description
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.

Classifications	Description
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 4.8.4

4.8.4 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Phoenix, Pig, Presto-Sandbox, Spark, Sqoop-Sandbox, Tez, Zeppelin-Sandbox, and ZooKeeper-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.8.4	emr-4.8.3	emr-4.8.2	emr-4.8.1
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-

	emr-4.8.4	emr-4.8.3	emr-4.8.2	emr-4.8.1
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.152.3	0.151
Spark	1.6.3	1.6.3	1.6.2	1.6.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-

	emr-4.8.4	emr-4.8.3	emr-4.8.2	emr-4.8.1
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.8	3.4.8

4.8.4 release notes

The following release notes include information for the Amazon EMR 4.8.4 release. Changes are relative to the Amazon EMR 4.8.3 release.

Release date: February 7, 2017

Minor changes, bug fixes, and enhancements were made in this release.

4.8.4 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form `*CommunityVersion*-amzn-*EmrVersion*`. The `*EmrVersion*` starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.2.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.2.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.14.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS node-level service for storing blocks.

Component	Version	Description
hadoop-hdfs-library	2.7.3-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.2	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.2	Service for serving one or more HBase regions.

Component	Version	Description
hbase-client	1.2.2	HBase command-line client.
hbase-rest-server	1.2.2	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.2	Service providing a Thrift endpoint to HBase.
hcatalog-client	1.0.0-amzn-8	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	1.0.0-amzn-8	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	1.0.0-amzn-8	HTTP endpoint providing a REST interface to HCatalog.
hive-client	1.0.0-amzn-8	Hive command line client.
hive-metastore-server	1.0.0-amzn-8	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	1.0.0-amzn-8	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-7	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.

Component	Version	Description
mysql-server	5.5.54+	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.157.1	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.157.1	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.6.3	Spark command-line clients.
spark-history-server	1.6.3	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.6.3	In-memory execution engine for YARN.
spark-yarn-slave	1.6.3	Apache Spark libraries needed by YARN slaves.

Component	Version	Description
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.25+	Apache HTTP server.
zeppelin-server	0.6.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.9	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.9	ZooKeeper command line client.

4.8.4 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-4.8.4 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.

Classifications	Description
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.

Classifications	Description
hcatalog-webhcat-log4j	Change values in HCatalog WebHCat's log4j.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-env	Change values in the Hive environment.
hive-exec-log4j	Change values in Hive's hive-exec-log4j.pr operties file.
hive-log4j	Change values in Hive's hive-log4j.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.

Classifications	Description
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.

Classifications	Description
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.

Classifications	Description
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 4.8.3

4.8.3 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Phoenix, Pig, Presto-Sandbox, Spark, Sqoop-Sandbox, Tez, Zeppelin-Sandbox, and ZooKeeper-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.8.3	emr-4.8.2	emr-4.8.1	emr-4.8.0
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-4.8.3	emr-4.8.2	emr-4.8.1	emr-4.8.0
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.152.3	0.151	0.151
Spark	1.6.3	1.6.2	1.6.2	1.6.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.8	3.4.8	3.4.8

4.8.3 release notes

The following release notes include information for the Amazon EMR 4.8.3 release. Changes are relative to the Amazon EMR 4.8.2 release.

Release date: December 29, 2016

Upgrades

- Upgraded to Presto 0.157.1. For more information, see Presto Release Notes in the Presto documentation.
- Upgraded to Spark 1.6.3. For more information, see Spark Release Notes in the Apache Spark documentation.
- Upgraded to ZooKeeper 3.4.9. For more information, see ZooKeeper Release Notes in the Apache ZooKeeper documentation.

Changes and enhancements

- Added support for the Amazon EC2 m4.16xlarge instance type in Amazon EMR version 4.8.3 and later, excluding 5.0.0, 5.0.3, and 5.2.0.
- Amazon EMR releases are now based on Amazon Linux 2016.09. For more information, see https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/.

Known issues resolved from previous releases

- Fixed an issue in Hadoop where the ReplicationMonitor thread could get stuck for a long time because of a race between replication and deletion of the same file in a large cluster.
- Fixed an issue where ControlledJob#toString failed with a null pointer exception (NPE) when job status was not successfully updated.

4.8.3 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
emr-ddb	4.2.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.2.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.13.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.

Component	Version	Description
hadoop-client	2.7.3-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service for retrieving current and historical information for YARN applications.

Component	Version	Description
hbase-hmaster	1.2.2	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.2	Service for serving one or more HBase regions.
hbase-client	1.2.2	HBase command-line client.
hbase-rest-server	1.2.2	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.2	Service providing a Thrift endpoint to HBase.
hcatalog-client	1.0.0-amzn-8	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	1.0.0-amzn-8	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	1.0.0-amzn-8	HTTP endpoint providing a REST interface to HCatalog.
hive-client	1.0.0-amzn-8	Hive command line client.
hive-metastore-server	1.0.0-amzn-8	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.

Component	Version	Description
hive-server	1.0.0-amzn-8	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-7	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.52	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.157.1	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.157.1	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.6.3	Spark command-line clients.

Component	Version	Description
spark-history-server	1.6.3	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.6.3	In-memory execution engine for YARN.
spark-yarn-slave	1.6.3	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.23	Apache HTTP server.
zeppelin-server	0.6.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.9	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.9	ZooKeeper command line client.

4.8.3 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-4.8.3 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.

Classifications	Description
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j	Change values in HCatalog WebHCat's log4j.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-env	Change values in the Hive environment.
hive-exec-log4j	Change values in Hive's hive-exec-log4j.pr operties file.
hive-log4j	Change values in Hive's hive-log4j.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.

Classifications	Description
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.

Classifications	Description
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.

Classifications	Description
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 4.8.2

4.8.2 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Phoenix, Pig, Presto-Sandbox, Spark, Sqoop-Sandbox, Tez, Zeppelin-Sandbox, and ZooKeeper-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases

- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.8.2	emr-4.8.1	emr-4.8.0	emr-4.7.4
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.1
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-

	emr-4.8.2	emr-4.8.1	emr-4.8.0	emr-4.7.4
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.152.3	0.151	0.151	0.148
Spark	1.6.2	1.6.2	1.6.2	1.6.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.3
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.5.6
ZooKeeper	-	-	-	-

	emr-4.8.2	emr-4.8.1	emr-4.8.0	emr-4.7.4
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

4.8.2 release notes

The following release notes include information for the Amazon EMR 4.8.2 release. Changes are relative to the Amazon EMR 4.8.0 release.

Release date: October 24, 2016

Upgrades

- Upgraded to Hadoop 2.7.3
- Upgraded to Presto 0.152.3, which includes support for the Presto web interface. You can access the Presto web interface on the Presto coordinator using port 8889. For more information about the Presto web interface, see Web Interface in the Presto documentation.
- Amazon EMR releases are now based on Amazon Linux 2016.09. For more information, see https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/.

4.8.2 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	4.1.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.1.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.10.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.3-amzn-0	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFS node-level service for storing blocks.

Component	Version	Description
hadoop-hdfs-library	2.7.3-amzn-0	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.3-amzn-0	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.3-amzn-0	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.3-amzn-0	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.3-amzn-0	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.2	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.2	Service for serving one or more HBase regions.

Component	Version	Description
hbase-client	1.2.2	HBase command-line client.
hbase-rest-server	1.2.2	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.2	Service providing a Thrift endpoint to HBase.
hcatalog-client	1.0.0-amzn-7	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	1.0.0-amzn-7	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	1.0.0-amzn-7	HTTP endpoint providing a REST interface to HCatalog.
hive-client	1.0.0-amzn-7	Hive command line client.
hive-metastore-server	1.0.0-amzn-7	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	1.0.0-amzn-7	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-7	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.

Component	Version	Description
mysql-server	5.5.52	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.152.3	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.152.3	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.6.2	Spark command-line clients.
spark-history-server	1.6.2	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.6.2	In-memory execution engine for YARN.
spark-yarn-slave	1.6.2	Apache Spark libraries needed by YARN slaves.

Component	Version	Description
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.23	Apache HTTP server.
zeppelin-server	0.6.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.8	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.8	ZooKeeper command line client.

4.8.2 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-4.8.2 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.

Classifications	Description
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.

Classifications	Description
hcatalog-webhcat-log4j	Change values in HCatalog WebHCat's log4j.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-env	Change values in the Hive environment.
hive-exec-log4j	Change values in Hive's hive-exec-log4j.pr operties file.
hive-log4j	Change values in Hive's hive-log4j.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.

Classifications	Description
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.

Classifications	Description
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.propert ies file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.

Classifications	Description
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 4.8.1

4.8.1 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Phoenix, Pig, Presto-Sandbox, Spark, Sqoop-Sandbox, Tez, Zeppelin-Sandbox, and ZooKeeper-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.8.1	emr-4.8.0	emr-4.7.4	emr-4.7.3
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.1	1.2.1
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-4.8.1	emr-4.8.0	emr-4.7.4	emr-4.7.3
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.151	0.151	0.148	0.148
Spark	1.6.2	1.6.2	1.6.2	1.6.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.3	0.8.3
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.5.6	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

4.8.1 release notes

4.8.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	3.2.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.1.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.9.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati

Component	Version	Description
		ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.2-amzn-4	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-4	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.2-amzn-4	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.2-amzn-4	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.2-amzn-4	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.2-amzn-4	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.2-amzn-4	MapReduce execution engine libraries for running a MapReduce application.

Component	Version	Description
hadoop-yarn-nodemanager	2.7.2-amzn-4	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.2-amzn-4	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.2-amzn-4	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.2	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.2	Service for serving one or more HBase regions.
hbase-client	1.2.2	HBase command-line client.
hbase-rest-server	1.2.2	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.2	Service providing a Thrift endpoint to HBase.
hcatalog-client	1.0.0-amzn-7	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	1.0.0-amzn-7	Service providing HCatalog, a table and storage management layer for distributed applications.

Component	Version	Description
hcatalog-webhcat-server	1.0.0-amzn-7	HTTP endpoint providing a REST interface to HCatalog.
hive-client	1.0.0-amzn-7	Hive command line client.
hive-metastore-server	1.0.0-amzn-7	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	1.0.0-amzn-7	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-7	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.51	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API

Component	Version	Description
presto-coordinator	0.151	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.151	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.6.2	Spark command-line clients.
spark-history-server	1.6.2	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.6.2	In-memory execution engine for YARN.
spark-yarn-slave	1.6.2	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.23	Apache HTTP server.
zeppelin-server	0.6.1	Web-based notebook that enables interactive data analytics.

Component	Version	Description
zookeeper-server	3.4.8	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.8	ZooKeeper command line client.

4.8.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-4.8.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration

Classifications	Description
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j	Change values in HCatalog WebHCat's log4j.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-env	Change values in the Hive environment.
hive-exec-log4j	Change values in Hive's hive-exec-log4j.pr operties file.
hive-log4j	Change values in Hive's hive-log4j.properties file.

Classifications	Description
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.

Classifications	Description
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.
presto-connector-postgresql	Change values in Presto's postgresql.properties file.

Classifications	Description
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 4.8.0

4.8.0 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Phoenix, Pig, Presto-Sandbox, Spark, Sqoop-Sandbox, Tez, Zeppelin-Sandbox, and ZooKeeper-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.8.0	emr-4.7.4	emr-4.7.3	emr-4.7.2
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.1	1.2.1	1.2.1

	emr-4.8.0	emr-4.7.4	emr-4.7.3	emr-4.7.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.151	0.148	0.148	0.148
Spark	1.6.2	1.6.2	1.6.2	1.6.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6

	emr-4.8.0	emr-4.7.4	emr-4.7.3	emr-4.7.2
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.3	0.8.3	0.8.3
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.5.6	0.5.6	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

4.8.0 release notes

The following release notes include information for the Amazon EMR 4.8.0 release. Changes are relative to the Amazon EMR 4.7.2 release.

Release date: September 7, 2016

Upgrades

- Upgraded to HBase 1.2.2
- Upgraded to Presto-Sandbox 0.151
- Upgraded to Tez 0.8.4
- Upgraded to Zeppelin-Sandbox 0.6.1

Changes and enhancements

- Fixed an issue in YARN where the ApplicationMaster would attempt to clean up containers that no longer exist because their instances have been terminated.
- Corrected the hive-server2 URL for Hive2 actions in the Oozie examples.

- Added support for additional Presto catalogs.
- Backported patches: HIVE-8948, HIVE-12679, HIVE-13405, PHOENIX-3116, HADOOP-12689
- Added support for security configurations, which allow you to create and apply encryption options more easily. For more information, see Data Encryption.

4.8.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	3.2.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.1.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.

Component	Version	Description
emrfs	2.9.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.2-amzn-4	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-4	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.2-amzn-4	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.2-amzn-4	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.2-amzn-4	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.2-amzn-4	Cryptographic key management server based on Hadoop's KeyProvider API.

Component	Version	Description
hadoop-mapred	2.7.2-amzn-4	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.2-amzn-4	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.2-amzn-4	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.2-amzn-4	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.2	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.2	Service for serving one or more HBase regions.
hbase-client	1.2.2	HBase command-line client.
hbase-rest-server	1.2.2	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.2	Service providing a Thrift endpoint to HBase.
hcatalog-client	1.0.0-amzn-7	The 'hcat' command line client for manipulating hcatalog-server.

Component	Version	Description
hcatalog-server	1.0.0-amzn-7	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	1.0.0-amzn-7	HTTP endpoint providing a REST interface to HCatalog.
hive-client	1.0.0-amzn-7	Hive command line client.
hive-metastore-server	1.0.0-amzn-7	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	1.0.0-amzn-7	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-7	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.51	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client

Component	Version	Description
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.151	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.151	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.6.2	Spark command-line clients.
spark-history-server	1.6.2	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.6.2	In-memory execution engine for YARN.
spark-yarn-slave	1.6.2	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.4	The tez YARN application and libraries.
webserver	2.4.23	Apache HTTP server.

Component	Version	Description
zeppelin-server	0.6.1	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.8	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.8	ZooKeeper command line client.

4.8.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see [Configure applications](#).

emr-4.8.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.

Classifications	Description
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j	Change values in HCatalog WebHCat's log4j.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-env	Change values in the Hive environment.

Classifications	Description
hive-exec-log4j	Change values in Hive's hive-exec-log4j.pr operties file.
hive-log4j	Change values in Hive's hive-log4j.properties file.
hive-site	Change values in Hive's hive-site.xml file
hiveserver2-site	Change values in Hive Server2's hiveserver2-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.

Classifications	Description
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-blackhole	Change values in Presto's blackhole.properties file.
presto-connector-cassandra	Change values in Presto's cassandra.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
presto-connector-jmx	Change values in Presto's jmx.properties file.
presto-connector-kafka	Change values in Presto's kafka.properties file.
presto-connector-localfile	Change values in Presto's localfile.properties file.
presto-connector-mongodb	Change values in Presto's mongodb.properties file.
presto-connector-mysql	Change values in Presto's mysql.properties file.

Classifications	Description
presto-connector-postgresql	Change values in Presto's postgresql.properties file.
presto-connector-raptor	Change values in Presto's raptor.properties file.
presto-connector-redis	Change values in Presto's redis.properties file.
presto-connector-tpch	Change values in Presto's tpch.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.

Classifications	Description
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 4.7.4

4.7.4 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Phoenix, Pig, Presto-Sandbox, Spark, Sqoop-Sandbox, Tez, Zeppelin-Sandbox, and ZooKeeper-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.7.4	emr-4.7.3	emr-4.7.2	emr-4.7.1
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-

	emr-4.7.4	emr-4.7.3	emr-4.7.2	emr-4.7.1
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.1	1.2.1	1.2.1
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-

	emr-4.7.4	emr-4.7.3	emr-4.7.2	emr-4.7.1
Presto-Sandbox	0.148	0.148	0.148	0.147
Spark	1.6.2	1.6.2	1.6.2	1.6.1
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.3	0.8.3	0.8.3	0.8.3
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

4.7.4 release notes

This is a patch release to add AWS Signature Version 4 authentication for requests to Amazon S3. All applications and components are the same as the previous Amazon EMR release.

> **Important**
>
> In this release version, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. For more information, see Whats New.

4.7.4 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	3.2.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.1.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.8.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.

Component	Version	Description
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.2-amzn-3	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.2-amzn-3	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.2-amzn-3	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.2-amzn-3	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.2-amzn-3	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARN service for managing containers on an individual node.

Component	Version	Description
hadoop-yarn-resour cemanager	2.7.2-amzn-3	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.1	Service for serving one or more HBase regions.
hbase-client	1.2.1	HBase command-line client.
hbase-rest-server	1.2.1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.1	Service providing a Thrift endpoint to HBase.
hcatalog-client	1.0.0-amzn-6	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	1.0.0-amzn-6	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	1.0.0-amzn-6	HTTP endpoint providing a REST interface to HCatalog.

Component	Version	Description
hive-client	1.0.0-amzn-6	Hive command line client.
hive-metastore-server	1.0.0-amzn-6	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	1.0.0-amzn-6	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-7	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.46	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.148	Service for accepting queries and managing query execution among presto-wo rkers.

Component	Version	Description
presto-worker	0.148	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.6.2	Spark command-line clients.
spark-history-server	1.6.2	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.6.2	In-memory execution engine for YARN.
spark-yarn-slave	1.6.2	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.3	The tez YARN application and libraries.
webserver	2.4.23	Apache HTTP server.
zeppelin-server	0.5.6-incubating	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.8	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.

Component	Version	Description
zookeeper-client	3.4.8	ZooKeeper command line client.

4.7.4 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-4.7.4 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.

Classifications	Description
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j	Change values in HCatalog WebHCat's log4j.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-env	Change values in the Hive environment.
hive-exec-log4j	Change values in Hive's hive-exec-log4j.properties file.
hive-log4j	Change values in Hive's hive-log4j.properties file.
hive-site	Change values in Hive's hive-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.

Classifications	Description
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.

Classifications	Description
presto-config	Change values in Presto's config.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 4.7.3

4.7.3 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Phoenix, Pig, Presto-Sandbox, Spark, Sqoop-Sandbox, Tez, Zeppelin-Sandbox, and ZooKeeper-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.7.3	emr-4.7.2	emr-4.7.1	emr-4.7.0
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.1	1.2.1	1.2.1

	emr-4.7.3	emr-4.7.2	emr-4.7.1	emr-4.7.0
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.0	0.12.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.148	0.148	0.147	0.147
Spark	1.6.2	1.6.2	1.6.1	1.6.1
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6

	emr-4.7.3	emr-4.7.2	emr-4.7.1	emr-4.7.0
TensorFlow	-	-	-	-
Tez	0.8.3	0.8.3	0.8.3	0.8.3
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

4.7.3 release notes

4.7.3 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	3.2.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.1.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.8.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.2-amzn-3	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFS node-level service for storing blocks.

Component	Version	Description
hadoop-hdfs-library	2.7.2-amzn-3	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.2-amzn-3	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.2-amzn-3	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.2-amzn-3	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.2-amzn-3	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.1	Service for serving one or more HBase regions.

Component	Version	Description
hbase-client	1.2.1	HBase command-line client.
hbase-rest-server	1.2.1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.1	Service providing a Thrift endpoint to HBase.
hcatalog-client	1.0.0-amzn-6	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	1.0.0-amzn-6	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	1.0.0-amzn-6	HTTP endpoint providing a REST interface to HCatalog.
hive-client	1.0.0-amzn-6	Hive command line client.
hive-metastore-server	1.0.0-amzn-6	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	1.0.0-amzn-6	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-7	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.2	Library for machine learning.

Component	Version	Description
mysql-server	5.5.46	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.148	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.148	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.6.2	Spark command-line clients.
spark-history-server	1.6.2	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.6.2	In-memory execution engine for YARN.
spark-yarn-slave	1.6.2	Apache Spark libraries needed by YARN slaves.

Component	Version	Description
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.3	The tez YARN application and libraries.
webserver	2.4.23	Apache HTTP server.
zeppelin-server	0.5.6-incubating	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.8	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.8	ZooKeeper command line client.

4.7.3 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-4.7.3 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.

Classifications	Description
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.

Classifications	Description
hcatalog-webhcat-log4j	Change values in HCatalog WebHCat's log4j.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-env	Change values in the Hive environment.
hive-exec-log4j	Change values in Hive's hive-exec-log4j.pr operties file.
hive-log4j	Change values in Hive's hive-log4j.properties file.
hive-site	Change values in Hive's hive-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.

Classifications	Description
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.

Classifications	Description
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 4.7.2

4.7.2 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Phoenix, Pig, Presto-Sandbox, Spark, Sqoop-Sandbox, Tez, Zeppelin-Sandbox, and ZooKeeper-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases

- [Application versions in Amazon EMR 4.x releases]

Application version information

	emr-4.7.2	emr-4.7.1	emr-4.7.0	emr-4.6.1
AWS SDK for Java	1.10.75	1.10.75	1.10.75	1.10.27
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.1	1.2.1	1.2.0
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-

	emr-4.7.2	emr-4.7.1	emr-4.7.0	emr-4.6.1
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.0	0.12.0	0.11.1
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	-
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.148	0.147	0.147	0.143
Spark	1.6.2	1.6.1	1.6.1	1.6.1
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.3	0.8.3	0.8.3	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

4.7.2 release notes

The following release notes include information for Amazon EMR 4.7.2.

Release date: July 15, 2016

Features

- Upgraded to Mahout 0.12.2
- Upgraded to Presto 0.148
- Upgraded to Spark 1.6.2
- You can now create an AWSCredentialsProvider for use with EMRFS using a URI as a parameter. For more information, see Create an AWSCredentialsProvider for EMRFS.
- EMRFS now allows users to configure a custom DynamoDB endpoint for their Consistent View metadata using the `fs.s3.consistent.dynamodb.endpoint` property in `emrfs-site.xml`.
- Added a script in `/usr/bin` called `spark-example`, which wraps `/usr/lib/spark/spark/bin/run-example` so you can run examples directly. For instance, to run the SparkPi example that comes with the Spark distribution, you can run `spark-example SparkPi 100` from the command line or using `command-runner.jar` as a step in the API.

Known issues resolved from previous releases

- Fixed an issue where Oozie had the `spark-assembly.jar` was not in the correct location when Spark was also installed, which resulted in failure to launch Spark applications with Oozie.
- Fixed an issue with Spark Log4j-based logging in YARN containers.

4.7.2 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2

has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	3.2.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.1.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.8.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.2-amzn-3	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.

Component	Version	Description
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.2-amzn-3	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.2-amzn-3	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.2-amzn-3	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.2-amzn-3	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.2-amzn-3	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.

Component	Version	Description
hbase-region-server	1.2.1	Service for serving one or more HBase regions.
hbase-client	1.2.1	HBase command-line client.
hbase-rest-server	1.2.1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.1	Service providing a Thrift endpoint to HBase.
hcatalog-client	1.0.0-amzn-6	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	1.0.0-amzn-6	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	1.0.0-amzn-6	HTTP endpoint providing a REST interface to HCatalog.
hive-client	1.0.0-amzn-6	Hive command line client.
hive-metastore-server	1.0.0-amzn-6	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	1.0.0-amzn-6	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-7	Web application for analyzing data using Hadoop ecosystem applications

Component	Version	Description
mahout-client	0.12.2	Library for machine learning.
mysql-server	5.5.46	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.148	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.148	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.6.2	Spark command-line clients.
spark-history-server	1.6.2	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.6.2	In-memory execution engine for YARN.

Component	Version	Description
spark-yarn-slave	1.6.2	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.3	The tez YARN application and libraries.
webserver	2.4.23	Apache HTTP server.
zeppelin-server	0.5.6-incubating	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.8	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.8	ZooKeeper command line client.

4.7.2 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-4.7.2 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.

Classifications	Description
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hadoop-ssl-server	Change hadoop ssl server configuration
hadoop-ssl-client	Change hadoop ssl client configuration
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.

Classifications	Description
hcatalog-webhcat-log4j	Change values in HCatalog WebHCat's log4j.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-env	Change values in the Hive environment.
hive-exec-log4j	Change values in Hive's hive-exec-log4j.properties file.
hive-log4j	Change values in Hive's hive-log4j.properties file.
hive-site	Change values in Hive's hive-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environment.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.

Classifications	Description
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.

Classifications	Description
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 4.7.1

4.7.1 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Phoenix, Pig, Presto-Sandbox, Spark, Sqoop-Sandbox, Tez, Zeppelin-Sandbox, and ZooKeeper-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases

- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.7.1	emr-4.7.0	emr-4.6.1	emr-4.6.0
AWS SDK for Java	1.10.75	1.10.75	1.10.27	1.10.27
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.1	1.2.0	1.2.0
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-

	emr-4.7.1	emr-4.7.0	emr-4.6.1	emr-4.6.0
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.0	0.12.0	0.11.1	0.11.1
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	-	-
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.147	0.147	0.143	0.143
Spark	1.6.1	1.6.1	1.6.1	1.6.1
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.3	0.8.3	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

4.7.1 release notes

The following release notes include information for Amazon EMR 4.7.1.

Release date: June 10, 2016

Known issues resolved from previous releases

- Fixed an issue that extended the startup time of clusters launched in a VPC with private subnets. The bug only impacted clusters launched with the Amazon EMR 4.7.0 release.
- Fixed an issue that improperly handled listing of files in Amazon EMR for clusters launched with the Amazon EMR 4.7.0 release.

4.7.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	3.1.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.0.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.

Component	Version	Description
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.7.1	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.2-amzn-2	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-2	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.2-amzn-2	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.2-amzn-2	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.2-amzn-2	HTTP endpoint for HDFS operations.

Component	Version	Description
hadoop-kms-server	2.7.2-amzn-2	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.2-amzn-2	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.2-amzn-2	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.2-amzn-2	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.2-amzn-2	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.1	Service for serving one or more HBase regions.
hbase-client	1.2.1	HBase command-line client.
hbase-rest-server	1.2.1	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.1	Service providing a Thrift endpoint to HBase.

Component	Version	Description
hcatalog-client	1.0.0-amzn-5	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	1.0.0-amzn-5	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	1.0.0-amzn-5	HTTP endpoint providing a REST interface to HCatalog.
hive-client	1.0.0-amzn-5	Hive command line client.
hive-metastore-server	1.0.0-amzn-5	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	1.0.0-amzn-5	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-7	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.0	Library for machine learning.
mysql-server	5.5.46	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.

Component	Version	Description
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.147	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.147	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.6.1	Spark command-line clients.
spark-history-server	1.6.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.6.1	In-memory execution engine for YARN.
spark-yarn-slave	1.6.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
tez-on-yarn	0.8.3	The tez YARN application and libraries.

Component	Version	Description
webserver	2.4.18	Apache HTTP server.
zeppelin-server	0.5.6-incubating	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.8	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.8	ZooKeeper command line client.

4.7.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-4.7.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.

Classifications	Description
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j	Change values in HCatalog WebHCat's log4j.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-env	Change values in the Hive environment.
hive-exec-log4j	Change values in Hive's hive-exec-log4j.pr operties file.

Classifications	Description
hive-log4j	Change values in Hive's hive-log4j.properties file.
hive-site	Change values in Hive's hive-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.

Classifications	Description
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
tez-site	Change values in Tez's tez-site.xml file.

Classifications	Description
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 4.7.0

4.7.0 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Phoenix, Pig, Presto-Sandbox, Spark, Sqoop-Sandbox, Tez, Zeppelin-Sandbox, and ZooKeeper-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.7.0	emr-4.6.1	emr-4.6.0	emr-4.5.0
AWS SDK for Java	1.10.75	1.10.27	1.10.27	1.10.27

	emr-4.7.0	emr-4.6.1	emr-4.6.0	emr-4.5.0
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.0	1.2.0	-
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.0	0.11.1	0.11.1	0.11.1
Oozie	-	-	-	-

	emr-4.7.0	emr-4.6.1	emr-4.6.0	emr-4.5.0
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	-	-	-
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.147	0.143	0.143	0.140
Spark	1.6.1	1.6.1	1.6.1	1.6.1
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.3	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	-

4.7.0 release notes

Important

Amazon EMR 4.7.0 is deprecated. Use Amazon EMR 4.7.1 or later instead.

Release date: June 2, 2016

Features

- Added Apache Phoenix 4.7.0
- Added Apache Tez 0.8.3
- Upgraded to HBase 1.2.1
- Upgraded to Mahout 0.12.0
- Upgraded to Presto 0.147
- Upgraded the AWS SDK for Java to 1.10.75
- The final flag was removed from the `mapreduce.cluster.local.dir` property in `mapred-site.xml` to allow users to run Pig in local mode.
- Amazon Redshift JDBC Drivers Available on Cluster

  Amazon Redshift JDBC drivers are now included at `/usr/share/aws/redshift/jdbc`. `/usr/share/aws/redshift/jdbc/RedshiftJDBC41.jar` is the JDBC 4.1-compatible Amazon Redshift driver and `/usr/share/aws/redshift/jdbc/RedshiftJDBC4.jar` is the JDBC 4.0-compatible Amazon Redshift driver. For more information, see Configure a JDBC Connection in the *Amazon Redshift Management Guide*.
- Java 8

  Except for Presto, OpenJDK 1.7 is the default JDK used for all applications. However, both OpenJDK 1.7 and 1.8 are installed. For information about how to set `JAVA_HOME` for applications, see Configuring Applications to Use Java 8.

Known issues resolved from previous releases

- Fixed a kernel issue that significantly affected performance on Throughput Optimized HDD (st1) EBS volumes for Amazon EMR in emr-4.6.0.
- Fixed an issue where a cluster would fail if any HDFS encryption zone were specified without choosing Hadoop as an application.
- Changed the default HDFS write policy from `RoundRobin` to `AvailableSpaceVolumeChoosingPolicy`. Some volumes were not properly utilized with the RoundRobin configuration, which resulted in failed core nodes and an unreliable HDFS.
- Fixed an issue with the EMRFS CLI, which would cause an exception when creating the default DynamoDB metadata table for consistent views.

- Fixed a deadlock issue in EMRFS that potentially occurred during multipart rename and copy operations.
- Fixed an issue with EMRFS that caused the CopyPart size default to be 5 MB. The default is now properly set at 128 MB.
- Fixed an issue with the Zeppelin upstart configuration that potentially prevented you from stopping the service.
- Fixed an issue with Spark and Zeppelin, which prevented you from using the `s3a://` URI scheme because `/usr/lib/hadoop/hadoop-aws.jar` was not properly loaded in their respective classpath.
- Backported HUE-2484.
- Backported a commit from Hue 3.9.0 (no JIRA exists) to fix an issue with the HBase browser sample.
- Backported HIVE-9073.

4.7.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	3.1.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.0.0	Extra convenience libraries for the Hadoop ecosystem.

Component	Version	Description
emr-kinesis	3.2.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.4.0	Distributed copy application optimized for Amazon S3.
emrfs	2.7.1	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.2-amzn-2	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-2	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.2-amzn-2	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.2-amzn-2	HDFS service for tracking file names and block locations.

Component	Version	Description
hadoop-httpfs-server	2.7.2-amzn-2	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.2-amzn-2	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.2-amzn-2	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.2-amzn-2	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.2-amzn-2	YARN service for allocatin g and managing cluster resources and distributed applications.
hadoop-yarn-timeline-server	2.7.2-amzn-2	Service for retrieving current and historical information for YARN applications.
hbase-hmaster	1.2.1	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.1	Service for serving one or more HBase regions.
hbase-client	1.2.1	HBase command-line client.
hbase-rest-server	1.2.1	Service providing a RESTful HTTP endpoint for HBase.

Component	Version	Description
hbase-thrift-server	1.2.1	Service providing a Thrift endpoint to HBase.
hcatalog-client	1.0.0-amzn-5	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	1.0.0-amzn-5	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	1.0.0-amzn-5	HTTP endpoint providing a REST interface to HCatalog.
hive-client	1.0.0-amzn-5	Hive command line client.
hive-metastore-server	1.0.0-amzn-5	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	1.0.0-amzn-5	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-7	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.12.0	Library for machine learning.
mysql-server	5.5.46	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.

Component	Version	Description
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
phoenix-library	4.7.0-HBase-1.2	The phoenix libraries for server and client
phoenix-query-server	4.7.0-HBase-1.2	A light weight server providing JDBC access as well as Protocol Buffers and JSON format access to the Avatica API
presto-coordinator	0.147	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.147	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.6.1	Spark command-line clients.
spark-history-server	1.6.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.6.1	In-memory execution engine for YARN.
spark-yarn-slave	1.6.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.

Component	Version	Description
tez-on-yarn	0.8.3	The tez YARN application and libraries.
webserver	2.4.18	Apache HTTP server.
zeppelin-server	0.5.6-incubating	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.8	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.8	ZooKeeper command line client.

4.7.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-4.7.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.

Classifications	Description
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j	Change values in HCatalog WebHCat's log4j.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-env	Change values in the Hive environment.

Classifications	Description
hive-exec-log4j	Change values in Hive's hive-exec-log4j.pr operties file.
hive-log4j	Change values in Hive's hive-log4j.properties file.
hive-site	Change values in Hive's hive-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
phoenix-hbase-metrics	Change values in Phoenix's hadoop-metrics2-hbase.properties file.

Classifications	Description
phoenix-hbase-site	Change values in Phoenix's hbase-site.xml file.
phoenix-log4j	Change values in Phoenix's log4j.properties file.
phoenix-metrics	Change values in Phoenix's hadoop-metrics2-phoenix.properties file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.

Classifications	Description
tez-site	Change values in Tez's tez-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 4.6.1

4.6.1 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Pig, Presto-Sandbox, Spark, Sqoop-Sandbox, Zeppelin-Sandbox, and ZooKeeper-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.6.1	emr-4.6.0	emr-4.5.0	emr-4.4.0
AWS SDK for Java	1.10.27	1.10.27	1.10.27	1.10.27
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.0	1.2.0	-	-
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.1
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-4.6.1	emr-4.6.0	emr-4.5.0	emr-4.4.0
Mahout	0.11.1	0.11.1	0.11.1	0.11.1
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	-	-	-	-
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.143	0.143	0.140	0.136
Spark	1.6.1	1.6.1	1.6.1	1.6.0
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	-	-

4.6.1 release notes

4.6.1 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form `CommunityVersion-amzn-EmrVersion`. The `EmrVersion` starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	3.0.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.0.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.1.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.3.0	Distributed copy application optimized for Amazon S3.
emrfs	2.6.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati

Component	Version	Description
		ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.2-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.2-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.2-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.2-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.2-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.2-amzn-1	MapReduce execution engine libraries for running a MapReduce application.

Component	Version	Description
hadoop-yarn-nodemanager	2.7.2-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.2-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hbase-hmaster	1.2.0	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.0	Service for serving one or more HBase regions.
hbase-client	1.2.0	HBase command-line client.
hbase-rest-server	1.2.0	Service providing a RESTful HTTP endpoint for HBase.
hbase-thrift-server	1.2.0	Service providing a Thrift endpoint to HBase.
hcatalog-client	1.0.0-amzn-4	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	1.0.0-amzn-4	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	1.0.0-amzn-4	HTTP endpoint providing a REST interface to HCatalog.

Component	Version	Description
hive-client	1.0.0-amzn-4	Hive command line client.
hive-metastore-server	1.0.0-amzn-4	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	1.0.0-amzn-4	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-6	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.11.1	Library for machine learning.
mysql-server	5.5	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
presto-coordinator	0.143	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.143	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.6.1	Spark command-line clients.

Component	Version	Description
spark-history-server	1.6.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.6.1	In-memory execution engine for YARN.
spark-yarn-slave	1.6.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
webserver	2.4	Apache HTTP server.
zeppelin-server	0.5.6-incubating	Web-based notebook that enables interactive data analytics.
zookeeper-server	3.4.8	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.8	ZooKeeper command line client.

4.6.1 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-4.6.1 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.
hbase-metrics	Change values in HBase's hadoop-metrics2-hb aase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.

Classifications	Description
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j	Change values in HCatalog WebHCat's log4j.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-env	Change values in the Hive environment.
hive-exec-log4j	Change values in Hive's hive-exec-log4j.pr operties file.
hive-log4j	Change values in Hive's hive-log4j.properties file.
hive-site	Change values in Hive's hive-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.

Classifications	Description
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.

Classifications	Description
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 4.6.0

4.6.0 application versions

The following applications are supported in this release: Ganglia, HBase, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Pig, Presto-Sandbox, Spark, Sqoop-Sandbox, Zeppelin-Sandbox, and ZooKeeper-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.6.0	emr-4.5.0	emr-4.4.0	emr-4.3.0
AWS SDK for Java	1.10.27	1.10.27	1.10.27	1.10.27
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.0	-	-	-
HCatalog	1.0.0	1.0.0	1.0.0	-
Hadoop	2.7.2	2.7.2	2.7.1	2.7.1
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-4.6.0	emr-4.5.0	emr-4.4.0	emr-4.3.0
Mahout	0.11.1	0.11.1	0.11.1	0.11.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	-	-	-	-
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.143	0.140	0.136	0.130
Spark	1.6.1	1.6.1	1.6.0	1.6.0
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.5
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	-	-	-

4.6.0 release notes

The following release notes include information for the Amazon EMR 4.6.0 release.

- Added HBase 1.2.0
- Added Zookeeper-Sandbox 3.4.8
- Upgraded to Presto-Sandbox 0.143
- Amazon EMR releases are now based on Amazon Linux 2016.03.0. For more information, see https://aws.amazon.com/amazon-linux-ami/2016.03-release-notes/.
- Issue Affecting Throughput Optimized HDD (st1) EBS Volume Types

  An issue in the Linux kernel versions 4.2 and above significantly affects performance on Throughput Optimized HDD (st1) EBS volumes for EMR. This release (emr-4.6.0) uses kernel version 4.4.5 and hence is impacted. Therefore, we recommend not using emr-4.6.0 if you want to use st1 EBS volumes. You can use emr-4.5.0 or prior Amazon EMR releases with st1 without impact. In addition, we provide the fix with future releases.
- Python Defaults

  Python 3.4 is now installed by default, but Python 2.7 remains the system default. You may configure Python 3.4 as the system default using either a bootstrap action; you can use the configuration API to set PYSPARK_PYTHON export to `/usr/bin/python3.4` in the `spark-env` classification to affect the Python version used by PySpark.
- Java 8

  Except for Presto, OpenJDK 1.7 is the default JDK used for all applications. However, both OpenJDK 1.7 and 1.8 are installed. For information about how to set `JAVA_HOME` for applications, see Configuring Applications to Use Java 8.

Known issues resolved from previous releases

- Fixed an issue where application provisioning would sometimes randomly fail due to a generated password.
- Previously, `mysqld` was installed on all nodes. Now, it is only installed on the master instance and only if the chosen application includes `mysql-server` as a component. Currently, the following applications include the `mysql-server` component: HCatalog, Hive, Hue, Presto-Sandbox, and Sqoop-Sandbox.

- Changed `yarn.scheduler.maximum-allocation-vcores` to 80 from the default of 32, which fixes an issue introduced in emr-4.4.0 that mainly occurs with Spark while using the `maximizeResourceAllocation` option in a cluster whose core instance type is one of a few large instance types that have the YARN vcores set higher than 32; namely c4.8xlarge, cc2.8xlarge, hs1.8xlarge, i2.8xlarge, m2.4xlarge, r3.8xlarge, d2.8xlarge, or m4.10xlarge were affected by this issue.
- s3-dist-cp now uses EMRFS for all Amazon S3 nominations and no longer stages to a temporary HDFS directory.
- Fixed an issue with exception handling for client-side encryption multipart uploads.
- Added an option to allow users to change the Amazon S3 storage class. By default this setting is `STANDARD`. The `emrfs-site` configuration classification setting is `fs.s3.storageClass` and the possible values are `STANDARD`, `STANDARD_IA`, and `REDUCED_REDUNDANCY`. For more information about storage classes, see Storage Classes in the *Amazon Simple Storage Service User Guide*.

4.6.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	3.0.0	Amazon DynamoDB connector for Hadoop ecosystem applications.

Component	Version	Description
emr-goodies	2.0.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.1.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.3.0	Distributed copy application optimized for Amazon S3.
emrfs	2.6.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.2-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.2-amzn-1	HDFS command-line client and library

Component	Version	Description
hadoop-hdfs-namenode	2.7.2-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.2-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.2-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.2-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.2-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.2-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hbase-hmaster	1.2.0	Service for an HBase cluster responsible for coordination of Regions and execution of administrative commands.
hbase-region-server	1.2.0	Service for serving one or more HBase regions.
hbase-client	1.2.0	HBase command-line client.
hbase-rest-server	1.2.0	Service providing a RESTful HTTP endpoint for HBase.

Component	Version	Description
hbase-thrift-server	1.2.0	Service providing a Thrift endpoint to HBase.
hcatalog-client	1.0.0-amzn-4	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	1.0.0-amzn-4	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	1.0.0-amzn-4	HTTP endpoint providing a REST interface to HCatalog.
hive-client	1.0.0-amzn-4	Hive command line client.
hive-metastore-server	1.0.0-amzn-4	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	1.0.0-amzn-4	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-6	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.11.1	Library for machine learning.
mysql-server	5.5	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.

Component	Version	Description
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
presto-coordinator	0.143	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.143	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.6.1	Spark command-line clients.
spark-history-server	1.6.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.6.1	In-memory execution engine for YARN.
spark-yarn-slave	1.6.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
webserver	2.4	Apache HTTP server.
zeppelin-server	0.5.6-incubating	Web-based notebook that enables interactive data analytics.

Component	Version	Description
zookeeper-server	3.4.8	Centralized service for maintaining configura tion information, naming, providing distributed synchronization, and providing group services.
zookeeper-client	3.4.8	ZooKeeper command line client.

4.6.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-4.6.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hbase-env	Change values in HBase's environment.
hbase-log4j	Change values in HBase's hbase-log4j.proper ties file.

Classifications	Description
hbase-metrics	Change values in HBase's hadoop-metrics2-hbaase.properties file.
hbase-policy	Change values in HBase's hbase-policy.xml file.
hbase-site	Change values in HBase's hbase-site.xml file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j	Change values in HCatalog WebHCat's log4j.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-env	Change values in the Hive environment.
hive-exec-log4j	Change values in Hive's hive-exec-log4j.properties file.
hive-log4j	Change values in Hive's hive-log4j.properties file.
hive-site	Change values in Hive's hive-site.xml file
hue-ini	Change values in Hue's ini file

Classifications	Description
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.

Classifications	Description
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.
zookeeper-config	Change values in ZooKeeper's zoo.cfg file.
zookeeper-log4j	Change values in ZooKeeper's log4j.properties file.

Amazon EMR release 4.5.0

4.5.0 application versions

The following applications are supported in this release: Ganglia, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Pig, Presto-Sandbox, Spark, Sqoop-Sandbox, and Zeppelin-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.5.0	emr-4.4.0	emr-4.3.0	emr-4.2.0
AWS SDK for Java	1.10.27	1.10.27	1.10.27	1.10.27
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.6.0
HBase	-	-	-	-
HCatalog	1.0.0	1.0.0	-	-
Hadoop	2.7.2	2.7.1	2.7.1	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-

	emr-4.5.0	emr-4.4.0	emr-4.3.0	emr-4.2.0
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.1	0.11.1	0.11.0	0.11.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	-	-	-	-
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.140	0.136	0.130	0.125
Spark	1.6.1	1.6.0	1.6.0	1.5.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-

	emr-4.5.0	emr-4.4.0	emr-4.3.0	emr-4.2.0
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.5	0.5.5
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

4.5.0 release notes

The following release notes include information for the Amazon EMR 4.5.0 release.

Release date: April 4, 2016

Features

- Upgraded to Spark 1.6.1
- Upgraded to Hadoop 2.7.2
- Upgraded to Presto 0.140
- Added AWS KMS support for Amazon S3 server-side encryption.

Known issues resolved from previous releases

- Fixed an issue where MySQL and Apache servers would not start after a node was rebooted.
- Fixed an issue where IMPORT did not work correctly with non-partitioned tables stored in Amazon S3
- Fixed an issue with Presto where it requires the staging directory to be `/mnt/tmp` rather than `/tmp` when writing to Hive tables.

4.5.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system

processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	3.0.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.0.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.1.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.2.0	Distributed copy application optimized for Amazon S3.
emrfs	2.5.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.

Component	Version	Description
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.2-amzn-0	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-0	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.2-amzn-0	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.2-amzn-0	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.2-amzn-0	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.2-amzn-0	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.2-amzn-0	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.2-amzn-0	YARN service for managing containers on an individual node.
hadoop-yarn-resourcemanager	2.7.2-amzn-0	YARN service for allocating and managing cluster resources and distributed applications.

Component	Version	Description
hcatalog-client	1.0.0-amzn-4	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	1.0.0-amzn-4	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	1.0.0-amzn-4	HTTP endpoint providing a REST interface to HCatalog.
hive-client	1.0.0-amzn-4	Hive command line client.
hive-metastore-server	1.0.0-amzn-4	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	1.0.0-amzn-4	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-5	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.11.1	Library for machine learning.
mysql-server	5.5	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.

Component	Version	Description
presto-coordinator	0.140	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.140	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.6.1	Spark command-line clients.
spark-history-server	1.6.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.6.1	In-memory execution engine for YARN.
spark-yarn-slave	1.6.1	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
webserver	2.4	Apache HTTP server.
zeppelin-server	0.5.6-incubating	Web-based notebook that enables interactive data analytics.

4.5.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-4.5.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j	Change values in HCatalog WebHCat's log4j.properties.
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-env	Change values in the Hive environment.
hive-exec-log4j	Change values in Hive's hive-exec-log4j.pr operties file.

Classifications	Description
hive-log4j	Change values in Hive's hive-log4j.properties file.
hive-site	Change values in Hive's hive-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.

Classifications	Description
presto-config	Change values in Presto's config.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-site.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.

Amazon EMR release 4.4.0

4.4.0 application versions

The following applications are supported in this release: Ganglia, HCatalog, Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Pig, Presto-Sandbox, Spark, Sqoop-Sandbox, and Zeppelin-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.4.0	emr-4.3.0	emr-4.2.0	emr-4.1.0
AWS SDK for Java	1.10.27	1.10.27	1.10.27	Not tracked
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.6.0	-
HBase	-	-	-	-
HCatalog	1.0.0	-	-	-
Hadoop	2.7.1	2.7.1	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-

	emr-4.4.0	emr-4.3.0	emr-4.2.0	emr-4.1.0
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.1	0.11.0	0.11.0	0.11.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.0.1
Phoenix	-	-	-	-
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.136	0.130	0.125	0.119
Spark	1.6.0	1.6.0	1.5.2	1.5.0
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	-	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-

	emr-4.4.0	emr-4.3.0	emr-4.2.0	emr-4.1.0
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.5	0.5.5	0.6.0-SNAPSHOT
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

4.4.0 release notes

The following release notes include information for the Amazon EMR 4.4.0 release.

Release date: March 14, 2016

Features

- Added HCatalog 1.0.0
- Added Sqoop-Sandbox 1.4.6
- Upgraded to Presto 0.136
- Upgraded to Zeppelin 0.5.6
- Upgraded to Mahout 0.11.1
- Enabled `dynamicResourceAllocation` by default.
- Added a table of all configuration classifications for the release. For more information, see the Configuration Classifications table in Configuring Applications.

Known issues resolved from previous releases

- Fixed an issue where the `maximizeResourceAllocation` setting would not reserve enough memory for YARN ApplicationMaster daemons.
- Fixed an issue encountered with a custom DNS. If any entries in `resolve.conf` precede the custom entries provided, then the custom entries are not resolvable. This behavior was affected by clusters in a VPC where the default VPC name server is inserted as the top entry in `resolve.conf`.

- Fixed an issue where the default Python moved to version 2.7 and boto was not installed for that version.
- Fixed an issue where YARN containers and Spark applications would generate a unique Ganglia round robin database (rrd) file, which resulted in the first disk attached to the instance filling up. Because of this fix, YARN container level metrics have been disabled and Spark application level metrics have been disabled.
- Fixed an issue in log pusher where it would delete all empty log folders. The effect was that the Hive CLI was not able to log because log pusher was removing the empty `user` folder under `/var/log/hive`.
- Fixed an issue affecting Hive imports, which affected partitioning and resulted in an error during import.
- Fixed an issue where EMRFS and s3-dist-cp did not properly handle bucket names that contain periods.
- Changed a behavior in EMRFS so that in versioning-enabled buckets the `_$folder$` marker file is not continuously created, which may contribute to improved performance for versioning-enabled buckets.
- Changed the behavior in EMRFS such that it does not use instruction files except for cases where client-side encryption is enabled. If you want to delete instruction files while using client-side encryption, you can set the emrfs-site.xml property, `fs.s3.cse.cryptoStorageMode.deleteInstructionFiles.enabled`, to true.
- Changed YARN log aggregation to retain logs at the aggregation destination for two days. The default destination is your cluster HDFS storage. If you want to change this duration, change the value of `yarn.log-aggregation.retain-seconds` using the `yarn-site` configuration classification when you create your cluster. As always, you can save your application logs to Amazon S3 using the `log-uri` parameter when you create your cluster.

Patches Applied

- HIVE-9655
- HIVE-9183
- HADOOP-12810

4.4.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	3.0.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.0.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.1.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.2.0	Distributed copy application optimized for Amazon S3.
emrfs	2.4.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.

Component	Version	Description
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.7.1-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.1-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.1-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.1-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.1-amzn-1	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.1-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.1-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.1-amzn-1	YARN service for managing containers on an individual node.

Component	Version	Description
hadoop-yarn-resour cemanager	2.7.1-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hcatalog-client	1.0.0-amzn-3	The 'hcat' command line client for manipulating hcatalog-server.
hcatalog-server	1.0.0-amzn-3	Service providing HCatalog, a table and storage management layer for distributed applications.
hcatalog-webhcat-server	1.0.0-amzn-3	HTTP endpoint providing a REST interface to HCatalog.
hive-client	1.0.0-amzn-3	Hive command line client.
hive-metastore-server	1.0.0-amzn-3	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	1.0.0-amzn-3	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-5	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.11.1	Library for machine learning.
mysql-server	5.5	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.

Component	Version	Description
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
presto-coordinator	0.136	Service for accepting queries and managing query execution among presto-workers.
presto-worker	0.136	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.6.0	Spark command-line clients.
spark-history-server	1.6.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.6.0	In-memory execution engine for YARN.
spark-yarn-slave	1.6.0	Apache Spark libraries needed by YARN slaves.
sqoop-client	1.4.6	Apache Sqoop command-line client.
webserver	2.4	Apache HTTP server.
zeppelin-server	0.5.6-incubating	Web-based notebook that enables interactive data analytics.

4.4.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-4.4.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hcatalog-env	Change values in HCatalog's environment.
hcatalog-server-jndi	Change values in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Change values in HCatalog's proto-hive-site.xml.
hcatalog-webhcat-env	Change values in HCatalog WebHCat's environment.
hcatalog-webhcat-log4j	Change values in HCatalog WebHCat's log4j.properties.

Classifications	Description
hcatalog-webhcat-site	Change values in HCatalog WebHCat's webhcat-site.xml file.
hive-env	Change values in the Hive environment.
hive-exec-log4j	Change values in Hive's hive-exec-log4j.pr operties file.
hive-log4j	Change values in Hive's hive-log4j.properties file.
hive-site	Change values in Hive's hive-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.

Classifications	Description
oozie-site	Change values in Oozie's oozie-site.xml file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
sqoop-env	Change values in Sqoop's environment.
sqoop-oraoop-site	Change values in Sqoop OraOop's oraoop-si te.xml file.
sqoop-site	Change values in Sqoop's sqoop-site.xml file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.

Amazon EMR release 4.3.0

4.3.0 application versions

The following applications are supported in this release: Ganglia, Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Pig, Presto-Sandbox, Spark, and Zeppelin-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
AWS SDK for Java	1.10.27	1.10.27	Not tracked	Not tracked
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.6.0	-	-
HBase	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
HCatalog	-	-	-	-
Hadoop	2.7.1	2.6.0	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	-
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.0	0.11.0	0.11.0	0.10.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.0.1	-
Phoenix	-	-	-	-
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.130	0.125	0.119	-
Spark	1.6.0	1.5.2	1.5.0	1.4.1
Sqoop	-	-	-	-
Sqoop-Sandbox	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.5	0.5.5	0.6.0-SNAPSHOT	-
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

4.3.0 release notes

The following release notes include information for the Amazon EMR 4.3.0 release.

Release date: January 19, 2016

Features

- Upgraded to Hadoop 2.7.1
- Upgraded to Spark 1.6.0
- Upgraded Ganglia to 3.7.2
- Upgraded Presto to 0.130
- Amazon EMR made some changes to `spark.dynamicAllocation.enabled` when it is set to true; it is false by default. When set to true, this affects the defaults set by the `maximizeResourceAllocation` setting:
  - If `spark.dynamicAllocation.enabled` is set to true, `spark.executor.instances` is not set by `maximizeResourceAllocation`.
  - The `spark.driver.memory` setting is now configured based on the instance types in the cluster in a similar way to how `spark.executors.memory` is set. However, because the Spark

driver application may run on either the master or one of the core instances (for example, in YARN client and cluster modes, respectively), the `spark.driver.memory` setting is set based on the instance type of the smaller instance type between these two instance groups.

- The `spark.default.parallelism` setting is now set at twice the number of CPU cores available for YARN containers. In previous releases, this was half that value.
- The calculations for the memory overhead reserved for Spark YARN processes were adjusted to be more accurate, resulting in a small increase in the total amount of memory available to Spark (that is, `spark.executor.memory`).

Known issues resolved from previous releases

- YARN log aggregation is now enabled by default.
- Fixed an issue where logs would not be pushed to Amazon S3 logs bucket for the cluster when YARN log aggregation was enabled.
- YARN container sizes now have a new minimum of 32 across all node types.
- Fixed an issue with Ganglia that caused excessive disk I/O on the primary node in large clusters.
- Fixed an issue that prevented applications logs from being pushed to Amazon S3 when a cluster is shutting down.
- Fixed an issue in EMRFS CLI that caused certain commands to fail.
- Fixed an issue with Zeppelin that prevented dependencies from being loaded in the underlying SparkContext.
- Fixed an issue that resulted from issuing a resize attempting to add instances.
- Fixed an issue in Hive where CREATE TABLE AS SELECT makes excessive list calls to Amazon S3.
- Fixed an issue where large clusters would not provision properly when Hue, Oozie, and Ganglia are installed.
- Fixed an issue in s3-dist-cp where it would return a zero exit code even if it failed with an error.

4.3.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named myapp-component with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as 2.2-amzn-2.

Component	Version	Description
emr-ddb	3.0.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.0.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.1.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.1.0	Distributed copy application optimized for Amazon S3.
emrfs	2.3.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
ganglia-monitor	3.7.2	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.7.2	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.7.1	Web application for viewing metrics collected by the Ganglia metadata collector.

Component	Version	Description
hadoop-client	2.7.1-amzn-0	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.7.1-amzn-0	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.7.1-amzn-0	HDFS command-line client and library
hadoop-hdfs-namenode	2.7.1-amzn-0	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.7.1-amzn-0	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.7.1-amzn-0	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.7.1-amzn-0	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.7.1-amzn-0	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.7.1-amzn-0	YARN service for allocatin g and managing cluster resources and distributed applications.
hive-client	1.0.0-amzn-2	Hive command line client.

Component	Version	Description
hive-metastore-server	1.0.0-amzn-2	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	1.0.0-amzn-2	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-5	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.11.0	Library for machine learning.
mysql-server	5.5	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
presto-coordinator	0.130	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.130	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.6.0	Spark command-line clients.
spark-history-server	1.6.0	Web UI for viewing logged events for the lifetime of a completed Spark application.

Component	Version	Description
spark-on-yarn	1.6.0	In-memory execution engine for YARN.
spark-yarn-slave	1.6.0	Apache Spark libraries needed by YARN slaves.
webserver	2.4	Apache HTTP server.
zeppelin-server	0.5.5-incubating-amzn-1	Web-based notebook that enables interactive data analytics.

4.3.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-4.3.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.

Classifications	Description
hive-env	Change values in the Hive environment.
hive-exec-log4j	Change values in Hive's hive-exec-log4j.pr operties file.
hive-log4j	Change values in Hive's hive-log4j.properties file.
hive-site	Change values in Hive's hive-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.

Classifications	Description
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.

Amazon EMR release 4.2.0

4.2.0 application versions

The following applications are supported in this release: Ganglia, Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Pig, Presto-Sandbox, Spark, and Zeppelin-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
AWS SDK for Java	1.10.27	1.10.27	Not tracked	Not tracked
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.6.0	-	-
HBase	-	-	-	-
HCatalog	-	-	-	-
Hadoop	2.7.1	2.6.0	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Hue	3.7.1	3.7.1	3.7.1	-
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.0	0.11.0	0.11.0	0.10.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.0.1	-
Phoenix	-	-	-	-
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.130	0.125	0.119	-
Spark	1.6.0	1.5.2	1.5.0	1.4.1
Sqoop	-	-	-	-
Sqoop-Sandbox	-	-	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.5	0.5.5	0.6.0-SNAPSHOT	-
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

4.2.0 release notes

The following release notes include information for the Amazon EMR 4.2.0 release.

Release date: November 18, 2015

Features

- Added Ganglia support
- Upgraded to Spark 1.5.2
- Upgraded to Presto 0.125
- Upgraded Oozie to 4.2.0
- Upgraded Zeppelin to 0.5.5
- Upgraded the AWS SDK for Java to 1.10.27

Known issues resolved from previous releases

- Fixed an issue with the EMRFS CLI where it did not use the default metadata table name.
- Fixed an issue encountered when using ORC-backed tables in Amazon S3.
- Fixed an issue encountered with a Python version mismatch in the Spark configuration.
- Fixed an issue when a YARN node status fails to report because of DNS issues for clusters in a VPC.
- Fixed an issue encountered when YARN decommissioned nodes, resulting in hung applications or the inability to schedule new applications.

- Fixed an issue encountered when clusters terminated with status TIMED_OUT_STARTING.
- Fixed an issue encountered when including the EMRFS Scala dependency in other builds. The Scala dependency has been removed.

4.2.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *CommunityVersion*-amzn-*EmrVersion*. The *EmrVersion* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	3.0.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.0.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.1.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.0.0	Distributed copy application optimized for Amazon S3.
emrfs	2.2.0	Amazon S3 connector for Hadoop ecosystem applicati ons.

Component	Version	Description
ganglia-monitor	3.6.0	Embedded Ganglia agent for Hadoop ecosystem applicati ons along with the Ganglia monitoring agent.
ganglia-metadata-collector	3.6.0	Ganglia metadata collector for aggregating metrics from Ganglia monitoring agents.
ganglia-web	3.5.10	Web application for viewing metrics collected by the Ganglia metadata collector.
hadoop-client	2.6.0-amzn-2	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.6.0-amzn-2	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.6.0-amzn-2	HDFS command-line client and library
hadoop-hdfs-namenode	2.6.0-amzn-2	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.6.0-amzn-2	HTTP endpoint for HDFS operations.
hadoop-kms-server	2.6.0-amzn-2	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.6.0-amzn-2	MapReduce execution engine libraries for running a MapReduce application.

Component	Version	Description
hadoop-yarn-nodemanager	2.6.0-amzn-2	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.6.0-amzn-2	YARN service for allocatin g and managing cluster resources and distributed applications.
hive-client	1.0.0-amzn-1	Hive command line client.
hive-metastore-server	1.0.0-amzn-1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	1.0.0-amzn-1	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-5	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.11.0	Library for machine learning.
mysql-server	5.5	MySQL database server.
oozie-client	4.2.0	Oozie command-line client.
oozie-server	4.2.0	Service for accepting Oozie workflow requests.
presto-coordinator	0.125	Service for accepting queries and managing query execution among presto-wo rkers.

Component	Version	Description
presto-worker	0.125	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.5.2	Spark command-line clients.
spark-history-server	1.5.2	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.5.2	In-memory execution engine for YARN.
spark-yarn-slave	1.5.2	Apache Spark libraries needed by YARN slaves.
webserver	2.4	Apache HTTP server.
zeppelin-server	0.5.5-incubating-amzn-0	Web-based notebook that enables interactive data analytics.

4.2.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-4.2.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.

Classifications	Description
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hive-env	Change values in the Hive environment.
hive-exec-log4j	Change values in Hive's hive-exec-log4j.pr operties file.
hive-log4j	Change values in Hive's hive-log4j.properties file.
hive-site	Change values in Hive's hive-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properti es file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.

Classifications	Description
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.proper ties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
presto-connector-hive	Change values in Presto's hive.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.
spark-metrics	Change values in Spark's metrics.properties file.
yarn-env	Change values in the YARN environment.

Classifications	Description
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.

Amazon EMR release 4.1.0

4.1.0 application versions

The following applications are supported in this release: Hadoop, Hive, Hue, Mahout, Oozie-Sandbox, Pig, Presto-Sandbox, Spark, and Zeppelin-Sandbox.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
AWS SDK for Java	1.10.27	1.10.27	Not tracked	Not tracked
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.6.0	-	-
HBase	-	-	-	-
HCatalog	-	-	-	-
Hadoop	2.7.1	2.6.0	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	-
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.0	0.11.0	0.11.0	0.10.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.0.1	-
Phoenix	-	-	-	-
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Presto-Sandbox	0.130	0.125	0.119	-
Spark	1.6.0	1.5.2	1.5.0	1.4.1
Sqoop	-	-	-	-
Sqoop-Sandbox	-	-	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.5	0.5.5	0.6.0-SNAPSHOT	-
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

4.1.0 release notes

4.1.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2

has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	3.0.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.0.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.1.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.0.0	Distributed copy application optimized for Amazon S3.
emrfs	2.1.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
hadoop-client	2.6.0-amzn-1	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.6.0-amzn-1	HDFS node-level service for storing blocks.
hadoop-hdfs-library	2.6.0-amzn-1	HDFS command-line client and library
hadoop-hdfs-namenode	2.6.0-amzn-1	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.6.0-amzn-1	HTTP endpoint for HDFS operations.

Component	Version	Description
hadoop-kms-server	2.6.0-amzn-1	Cryptographic key management server based on Hadoop's KeyProvider API.
hadoop-mapred	2.6.0-amzn-1	MapReduce execution engine libraries for running a MapReduce application.
hadoop-yarn-nodemanager	2.6.0-amzn-1	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.6.0-amzn-1	YARN service for allocatin g and managing cluster resources and distributed applications.
hive-client	1.0.0-amzn-1	Hive command line client.
hive-metastore-server	1.0.0-amzn-1	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	1.0.0-amzn-1	Service for accepting Hive queries as web requests.
hue-server	3.7.1-amzn-4	Web application for analyzing data using Hadoop ecosystem applications
mahout-client	0.11.0	Library for machine learning.
mysql-server	5.5	MySQL database server.
oozie-client	4.0.1	Oozie command-line client.

Component	Version	Description
oozie-server	4.0.1	Service for accepting Oozie workflow requests.
presto-coordinator	0.119	Service for accepting queries and managing query execution among presto-wo rkers.
presto-worker	0.119	Service for executing pieces of a query.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.5.0	Spark command-line clients.
spark-history-server	1.5.0	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.5.0	In-memory execution engine for YARN.
spark-yarn-slave	1.5.0	Apache Spark libraries needed by YARN slaves.
zeppelin-server	0.6.0-incubating-SNAPSHOT	Web-based notebook that enables interactive data analytics.

4.1.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see Configure applications.

emr-4.1.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hdfs-encryption-zones	Configure HDFS encryption zones.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hive-env	Change values in the Hive environment.
hive-exec-log4j	Change values in Hive's hive-exec-log4j.pr operties file.
hive-log4j	Change values in Hive's hive-log4j.properties file.
hive-site	Change values in Hive's hive-site.xml file
hue-ini	Change values in Hue's ini file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
hadoop-kms-acls	Change values in Hadoop's kms-acls.xml file.
hadoop-kms-env	Change values in the Hadoop KMS environme nt.

Classifications	Description
hadoop-kms-log4j	Change values in Hadoop's kms-log4j.properties file.
hadoop-kms-site	Change values in Hadoop's kms-site.xml file.
mapred-env	Change values in the MapReduce application's environment.
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
oozie-env	Change values in Oozie's environment.
oozie-log4j	Change values in Oozie's oozie-log4j.properties file.
oozie-site	Change values in Oozie's oozie-site.xml file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
presto-log	Change values in Presto's log.properties file.
presto-config	Change values in Presto's config.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.
yarn-env	Change values in the YARN environment.

Classifications	Description
yarn-site	Change values in YARN's yarn-site.xml file.
zeppelin-env	Change values in the Zeppelin environment.

Amazon EMR release 4.0.0

4.0.0 application versions

The following applications are supported in this release: Hadoop, Hive, Mahout, Pig, and Spark.

The table below lists the application versions available in this release of Amazon EMR and the application versions in the preceding three Amazon EMR releases (when applicable).

For a comprehensive history of application versions for each release of Amazon EMR, see the following topics:

- Application versions in Amazon EMR 7.x releases
- Application versions in Amazon EMR 6.x releases
- Application versions in Amazon EMR 5.x releases
- Application versions in Amazon EMR 4.x releases

Application version information

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
AWS SDK for Java	1.10.27	1.10.27	Not tracked	Not tracked
Python	Not tracked	Not tracked	Not tracked	Not tracked
Scala	Not tracked	Not tracked	Not tracked	Not tracked
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Flink	-	-	-	-
Ganglia	3.7.2	3.6.0	-	-
HBase	-	-	-	-
HCatalog	-	-	-	-
Hadoop	2.7.1	2.6.0	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	-
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.0	0.11.0	0.11.0	0.10.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.0.1	-
Phoenix	-	-	-	-
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.130	0.125	0.119	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Spark	1.6.0	1.5.2	1.5.0	1.4.1
Sqoop	-	-	-	-
Sqoop-Sandbox	-	-	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.5	0.5.5	0.6.0-SNAPSHOT	-
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

4.0.0 release notes

4.0.0 component versions

The components that Amazon EMR installs with this release are listed below. Some are installed as part of big-data application packages. Others are unique to Amazon EMR and installed for system processes and features. These typically start with `emr` or `aws`. Big-data application packages in the most recent Amazon EMR release are usually the latest version found in the community. We make community releases available in Amazon EMR as quickly as possible.

Some components in Amazon EMR differ from community versions. These components have a version label in the form *`CommunityVersion`*`-amzn-`*`EmrVersion`*. The *`EmrVersion`* starts at 0. For example, if open source community component named `myapp-component` with version 2.2 has been modified three times for inclusion in different Amazon EMR releases, its release version is listed as `2.2-amzn-2`.

Component	Version	Description
emr-ddb	3.0.0	Amazon DynamoDB connector for Hadoop ecosystem applications.
emr-goodies	2.0.0	Extra convenience libraries for the Hadoop ecosystem.
emr-kinesis	3.0.0	Amazon Kinesis connector for Hadoop ecosystem applicati ons.
emr-s3-dist-cp	2.0.0	Distributed copy application optimized for Amazon S3.
emrfs	2.0.0	Amazon S3 connector for Hadoop ecosystem applicati ons.
hadoop-client	2.6.0-amzn-0	Hadoop command-line clients such as 'hdfs', 'hadoop', or 'yarn'.
hadoop-hdfs-datanode	2.6.0-amzn-0	HDFS node-level service for storing blocks.
hadoop-hdfs-namenode	2.6.0-amzn-0	HDFS service for tracking file names and block locations.
hadoop-httpfs-server	2.6.0-amzn-0	HTTP endpoint for HDFS operations.
hadoop-mapred	2.6.0-amzn-0	MapReduce execution engine libraries for running a MapReduce application.

Component	Version	Description
hadoop-yarn-nodemanager	2.6.0-amzn-0	YARN service for managing containers on an individual node.
hadoop-yarn-resour cemanager	2.6.0-amzn-0	YARN service for allocatin g and managing cluster resources and distributed applications.
hive-client	1.0.0-amzn-0	Hive command line client.
hive-metastore-server	1.0.0-amzn-0	Service for accessing the Hive metastore, a semantic repository storing metadata for SQL on Hadoop operation s.
hive-server	1.0.0-amzn-0	Service for accepting Hive queries as web requests.
mahout-client	0.10.0	Library for machine learning.
mysql-server	5.5	MySQL database server.
pig-client	0.14.0-amzn-0	Pig command-line client.
spark-client	1.4.1	Spark command-line clients.
spark-history-server	1.4.1	Web UI for viewing logged events for the lifetime of a completed Spark application.
spark-on-yarn	1.4.1	In-memory execution engine for YARN.
spark-yarn-slave	1.4.1	Apache Spark libraries needed by YARN slaves.

4.0.0 configuration classifications

Configuration classifications allow you to customize applications. These often correspond to a configuration XML file for the application, such as `hive-site.xml`. For more information, see [Configure applications](#).

emr-4.0.0 classifications

Classifications	Description
capacity-scheduler	Change values in Hadoop's capacity-scheduler .xml file.
core-site	Change values in Hadoop's core-site.xml file.
emrfs-site	Change EMRFS settings.
hadoop-env	Change values in the Hadoop environment for all Hadoop components.
hadoop-log4j	Change values in Hadoop's log4j.properties file.
hdfs-site	Change values in HDFS's hdfs-site.xml.
hive-env	Change values in the Hive environment.
hive-exec-log4j	Change values in Hive's hive-exec-log4j.pr operties file.
hive-log4j	Change values in Hive's hive-log4j.properties file.
hive-site	Change values in Hive's hive-site.xml file
httpfs-env	Change values in the HTTPFS environment.
httpfs-site	Change values in Hadoop's httpfs-site.xml file.
mapred-env	Change values in the MapReduce application's environment.

Classifications	Description
mapred-site	Change values in the MapReduce application's mapred-site.xml file.
pig-properties	Change values in Pig's pig.properties file.
pig-log4j	Change values in Pig's log4j.properties file.
spark	Amazon EMR-curated settings for Apache Spark.
spark-defaults	Change values in Spark's spark-defaults.conf file.
spark-env	Change values in the Spark environment.
spark-log4j	Change values in Spark's log4j.properties file.
yarn-env	Change values in the YARN environment.
yarn-site	Change values in YARN's yarn-site.xml file.

Amazon EMR 2.x and 3.x AMI versions

Note

AWS is updating the TLS configuration for all AWS API endpoints to a minimum version of TLS 1.2. Amazon EMR releases 3.10 and lower only support TLS 1.0/1.1 connections. After December 4, 2023, you won't be able to create clusters with Amazon EMR 3.10 and lower. If you use Amazon EMR 3.10 or lower, we recommend that you immediately test and migrate your workloads to the latest Amazon EMR release. For more information, see the AWS Security Blog.

Amazon EMR 2.x and 3.x releases, called *AMI versions*, are made available for pre-existing solutions that require them for compatibility reasons. We do not recommend creating new clusters or new solutions with these release versions. They lack features of newer releases and include outdated application packages.

We recommend that you build solutions using the most recent Amazon EMR release version.

The scope of differences between the 2.x and 3.x series release versions and recent Amazon EMR release versions is significant. Those differences range from how you create and configure a cluster to the ports and directory structure of applications on the cluster.

This section attempts to cover the most significant differences for Amazon EMR, as well as specific application configuration and management differences. It is not comprehensive. If you create and use clusters in the 2.x or 3.x series, you may encounter differences not covered in this section.

Topics

Creating a cluster with earlier AMI versions of Amazon EMR

Amazon EMR 2.x and 3.x releases are referenced by *AMI version*. With Amazon EMR release 4.0.0 and later, releases are referenced by release version, using a *release label* such as `emr-5.11.0`. This change is most apparent when you create a cluster using the AWS CLI or programmatically.

When you use the AWS CLI to create a cluster using an AMI release version, use the `--ami-version` option, for example, `--ami-version 3.11.0`. Many options, features, and applications introduced in Amazon EMR 4.0.0 and later are not available when you specify an `--ami-version`. For more information, see create-cluster in the *AWS CLI Command Reference*.

The following example AWS CLI command launches a cluster using an AMI version.

Note

Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.11.0 \
--applications Name=Hue Name=Hive Name=Pig \
--use-default-roles --ec2-attributes KeyName=myKey \
--instance-groups InstanceGroupType=MASTER,InstanceCount=1,\
InstanceType=m3.xlarge InstanceGroupType=CORE,InstanceCount=2,\
InstanceType=m3.xlarge --bootstrap-actions Path=s3://elasticmapreduce/bootstrap-
actions/configure-hadoop,\
Name="Configuring infinite JVM reuse",Args=["-m","mapred.job.reuse.jvm.num.tasks=-1"]
```

Programmatically, all Amazon EMR release versions use the `RunJobFlowRequest` action in the EMR API to create clusters. The following example Java code creates a cluster using AMI release version 3.11.0.

```
RunJobFlowRequest request = new RunJobFlowRequest()
   .withName("AmiVersion Cluster")
   .withAmiVersion("3.11.0")
   .withInstances(new JobFlowInstancesConfig()
    .withEc2KeyName("myKeyPair")
    .withInstanceCount(1)
    .withKeepJobFlowAliveWhenNoSteps(true)
    .withMasterInstanceType("m3.xlarge")
    .withSlaveInstanceType("m3.xlarge");
```

The following `RunJobFlowRequest` call uses a release label instead:

```
RunJobFlowRequest request = new RunJobFlowRequest()
   .withName("ReleaseLabel Cluster")
   .withReleaseLabel("emr-7.2.0")
   .withInstances(new JobFlowInstancesConfig()
    .withEc2KeyName("myKeyPair")
    .withInstanceCount(1)
    .withKeepJobFlowAliveWhenNoSteps(true)
    .withMasterInstanceType("m3.xlarge")
    .withSlaveInstanceType("m3.xlarge");
```

Configuring cluster size

When your cluster runs, Hadoop determines the number of mapper and reducer tasks needed to process the data. Larger clusters should have more tasks for better resource use and shorter processing time. Typically, an EMR cluster remains the same size during the entire cluster; you set the number of tasks when you create the cluster. When you resize a running cluster, you can vary the processing during the cluster execution. Therefore, instead of using a fixed number of tasks, you can vary the number of tasks during the life of the cluster. There are two configuration options to help set the ideal number of tasks:

- `mapred.map.tasksperslot`
- `mapred.reduce.tasksperslot`

You can set both options in the `mapred-conf.xml` file. When you submit a job to the cluster, the job client checks the current total number of map and reduce slots available clusterwide. The job client then uses the following equations to set the number of tasks:

- `mapred.map.tasks` = `mapred.map.tasksperslot` * map slots in cluster
- `mapred.reduce.tasks` = `mapred.reduce.tasksperslot` * reduce slots in cluster

The job client only reads the `tasksperslot` parameter if the number of tasks is not configured. You can override the number of tasks at any time, either for all clusters via a bootstrap action or individually per job by adding a step to change the configuration.

Amazon EMR withstands task node failures and continues cluster execution even if a task node becomes unavailable. Amazon EMR automatically provisions additional task nodes to replace those that fail.

You can have a different number of task nodes for each cluster step. You can also add a step to a running cluster to modify the number of task nodes. Because all steps are guaranteed to run sequentially by default, you can specify the number of running task nodes for any step.

Installing applications with earlier AMI versions of Amazon EMR

When using an AMI version, applications are installed in any number of ways, including using the `NewSupportedProducts` parameter for the [RunJobFlow](#) action, using bootstrap actions, and using the [Step](#) action.

Customizing cluster and application configuration with earlier AMI versions of Amazon EMR

Amazon EMR release version 4.0.0 introduced a simplified method of configuring applications using configuration classifications. For more information, see Configure applications. When using an AMI version, you configure applications using bootstrap actions along with arguments that you pass. For example, the `configure-hadoop` and `configure-daemons` bootstrap actions set Hadoop and YARN–specific environment properties like `--namenode-heap-size`. In more recent versions, these are configured using the `hadoop-env` and `yarn-env` configuration classifications. For bootstrap actions that perform common configurations, see the emr-bootstrap-actions repository on Github.

The following tables map bootstrap actions to configuration classifications in more recent Amazon EMR release versions.

Hadoop

Affected application file name	AMI version bootstrap action	Configuration classification
`core-site.xml`	`configure-hadoop -c`	`core-site`
`log4j.properties`	`configure-hadoop -l`	`hadoop-log4j`
`hdfs-site.xml`	`configure-hadoop -s`	`hdfs-site`
n/a	n/a	`hdfs-encryption-zones`
`mapred-site.xml`	`configure-hadoop -m`	`mapred-site`
`yarn-site.xml`	`configure-hadoop -y`	`yarn-site`
`httpfs-site.xml`	`configure-hadoop -t`	`httpfs-site`
`capacity-scheduler.xml`	`configure-hadoop -z`	`capacity-scheduler`
`yarn-env.sh`	`configure-daemons --resourcemanager-opts`	`yarn-env`

Hive

Affected application file name	AMI version bootstrap action	Configuration classification
hive-env.sh	n/a	hive-env
hive-site.xml	hive-script --install -hive-site ${MY_HIVE _SITE_FILE}	hive-site
hive-exec-log4j.pr operties	n/a	hive-exec-log4j
hive-log4j.propert ies	n/a	hive-log4j

EMRFS

Affected application file name	AMI version bootstrap action	Configuration classification
emrfs-site.xml	configure-hadoop -e	emrfs-site
n/a	s3get -s s3://cust om-provider.jar -d / usr/share/aws/emr/ auxlib/	emrfs-site (with new setting fs.s3.cse .encryptionMateria lsProvider.uri)

For a list of all classifications, see Configure applications.

Application environment variables

When using an AMI version, a hadoop-user-env.sh script is used along with the configure-daemons bootstrap action to configure the Hadoop environment. The script includes the following actions:

```
#!/bin/bash
export HADOOP_USER_CLASSPATH_FIRST=true;
```

```
echo "HADOOP_CLASSPATH=/path/to/my.jar" >> /home/hadoop/conf/hadoop-user-env.sh
```

In Amazon EMR release 4.x, you do the same using the hadoop-env configuration classification, as shown in the following example:

```
[
      {
         "Classification":"hadoop-env",
         "Properties":{

         },
         "Configurations":[
            {
               "Classification":"export",
               "Properties":{
                  "HADOOP_USER_CLASSPATH_FIRST":"true",
                  "HADOOP_CLASSPATH":"/path/to/my.jar"
               }
            }
         ]
      }
   ]
```

As another example, using configure-daemons and passing --namenode-heap-size=2048 and --namenode-opts=-XX:GCTimeRatio=19 is equivalent to the following configuration classifications.

```
[
      {
         "Classification":"hadoop-env",
         "Properties":{

         },
         "Configurations":[
            {
               "Classification":"export",
               "Properties":{
                  "HADOOP_DATANODE_HEAPSIZE":  "2048",
            "HADOOP_NAMENODE_OPTS":  "-XX:GCTimeRatio=19"
               }
            }
         ]
```

```
        }
    ]
```

Other application environment variables are no longer defined in `/home/hadoop/.bashrc`. Instead, they are primarily set in `/etc/default` files per component or application, such as `/etc/default/hadoop`. Wrapper scripts in `/usr/bin/` installed by application RPMs may also set additional environment variables before involving the actual bin script.

Service ports

When using an AMI version, some services use custom ports.

Changes in port settings

Setting	AMI version 3.x	Open-source default
fs.default.name	hdfs://emrDeterminedIP:9000	default (hdfs://*emrDeterminedIP* :8020)
dfs.datanode.address	0.0.0.0:9200	default (0.0.0.0:50010)
dfs.datanode.http.address	0.0.0.0:9102	default (0.0.0.0:50075)
dfs.datanode.https.address	0.0.0.0:9402	default (0.0.0.0:50475)
dfs.datanode.ipc.address	0.0.0.0:9201	default (0.0.0.0:50020)
dfs.http.address	0.0.0.0:9101	default (0.0.0.0:50070)
dfs.https.address	0.0.0.0:9202	default (0.0.0.0:50470)
dfs.secondary.http.address	0.0.0.0:9104	default (0.0.0.0:50090)
yarn.nodemanager.address	0.0.0.0:9103	default (${yarn.nodemanager.hostname}:0)
yarn.nodemanager.localizer.address	0.0.0.0:9033	default (${yarn.nodemanager.hostname}:8040)
yarn.nodemanager.webapp.address	0.0.0.0:9035	default (${yarn.nodemanager.hostname}:8042)

Setting	AMI version 3.x	Open-source default
yarn.resourcemanager.address	*emrDeterminedIP* :9022	default (${yarn.resourcemanager.hostname}:8032)
yarn.resourcemanager.admin.address	*emrDeterminedIP* :9025	default (${yarn.resourcemanager.hostname}:8033)
yarn.resourcemanager.resource-tracker.address	*emrDeterminedIP* :9023	default (${yarn.resourcemanager.hostname}:8031)
yarn.resourcemanager.scheduler.address	*emrDeterminedIP* :9024	default (${yarn.resourcemanager.hostname}:8030)
yarn.resourcemanager.webapp.address	0.0.0.0:9026	default (${yarn.resourcemanager.hostname}:8088)
yarn.web-proxy.address	*emrDeterminedIP* :9046	default (no-value)
yarn.resourcemanager.hostname	0.0.0.0 (default)	*emrDeterminedIP*

Note

The *emrDeterminedIP* is an IP address that is generated by Amazon EMR.

Users

When using an AMI version, the user `hadoop` runs all processes and owns all files. In Amazon EMR release version 4.0.0 and later, users exist at the application and component level.

Installation sequence, installed artifacts, and log file locations

When using an AMI version, application artifacts and their configuration directories are installed in the `/home/hadoop/application` directory. For example, if you installed Hive, the directory would be `/home/hadoop/hive`. In Amazon EMR release 4.0.0 and later, application artifacts are installed in the `/usr/lib/application` directory. When using an AMI version, log files are found in various places. The table below lists locations.

Changes in log locations on Amazon S3

Daemon or application	Directory location
instance-state	node/*instance-id* /instance-state/
hadoop-hdfs-namenode	daemons/*instance-id* /hadoop-hadoop-namenode.log
hadoop-hdfs-datanode	daemons/*instance-id* /hadoop-hadoop-datanode.log
hadoop-yarn (ResourceManager)	daemons/*instance-id* /yarn-hadoop-resourcemanager
hadoop-yarn (Proxy Server)	daemons/*instance-id* /yarn-hadoop-proxyserver
mapred-historyserver	daemons/*instance-id* /
httpfs	daemons/*instance-id* /httpfs.log
hive-server	node/*instance-id* /hive-server/hive-server.log
hive-metastore	node/*instance-id* /apps/hive.log
Hive CLI	node/*instance-id* /apps/hive.log
YARN applications user logs and container logs	task-attempts/
Mahout	N/A
Pig	N/A
spark-historyserver	N/A
mapreduce job history files	jobs/

Command runner

When using an AMI version, many scripts or programs, like /home/hadoop/contrib/streaming/hadoop-streaming.jar, are not placed on the shell login path environment, so you need to specify the full path when you use a jar file such as command-runner.jar or script-runner.jar to execute the scripts. The command-runner.jar is located on the AMI so there is no need to know a full URI as was the case with script-runner.jar.

Replication factor

The replication factor lets you configure when to start a Hadoop JVM. You can start a new Hadoop JVM for every task, which provides better task isolation, or you can share JVMs between tasks, providing lower framework overhead. If you are processing many small files, it makes sense to reuse the JVM many times to amortize the cost of start-up. However, if each task takes a long time or processes a large amount of data, then you might choose to not reuse the JVM to ensure that all memory is freed for subsequent tasks. When using an AMI version, you can customize the replication factor using the configure-hadoop bootstrap action to set the mapred.job.reuse.jvm.num.tasks property.

The following example demonstrates setting the JVM reuse factor for infinite JVM reuse.

Note

Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.11.0 \
--applications Name=Hue Name=Hive Name=Pig \
--use-default-roles --ec2-attributes KeyName=myKey \
--instance-groups InstanceGroupType=MASTER,InstanceCount=1,InstanceType=m3.xlarge \
InstanceGroupType=CORE,InstanceCount=2,InstanceType=m3.xlarge \
--bootstrap-actions Path=s3://elasticmapreduce/bootstrap-actions/configure-hadoop,\
Name="Configuring infinite JVM reuse",Args=["-m","mapred.job.reuse.jvm.num.tasks=-1"]
```

Hive application specifics for earlier AMI versions of Amazon EMR

Log files

Using Amazon EMR AMI versions 2.x and 3.x, Hive logs are saved to `/mnt/var/log/apps/`. In order to support concurrent versions of Hive, the version of Hive that you run determines the log file name, as shown in the following table.

Hive version	Log file name
0.13.1	`hive.log` **Note** Beginning with this version, Amazon EMR uses an unversioned file name, `hive.log`. Minor versions share the same log location as the major version.
0.11.0	`hive_0110.log` **Note** Minor versions of Hive 0.11.0, such as 0.11.0.1, share the same log file location as Hive 0.11.0.
0.8.1	`hive_081.log` **Note** Minor versions of Hive 0.8.1, such as Hive 0.8.1.1, share the same log file location as Hive 0.8.1.
0.7.1	`hive_07_1.log`

Hive version	Log file name
	Note Minor versions of Hive 0.7.1, such as Hive 0.7.1.3 and Hive 0.7.1.4, share the same log file location as Hive 0.7.1.
0.7	`hive_07.log`
0.5	`hive_05.log`
0.4	`hive.log`

Split input functionality

To implement split input functionality using Hive versions earlier than 0.13.1 (Amazon EMR AMI versions earlier 3.11.0), use the following:

```
hive> set hive.input.format=org.apache.hadoop.hive.ql.io.HiveCombineSplitsInputFormat;
hive> set mapred.min.split.size=100000000;
```

This functionality was deprecated with Hive 0.13.1. To get the same split input format functionality in Amazon EMR AMI Version 3.11.0, use the following:

```
set hive.hadoop.supports.splittable.combineinputformat=true;
```

Thrift service ports

Thrift is an RPC framework that defines a compact binary serialization format used to persist data structures for later analysis. Normally, Hive configures the server to operate on the following ports.

Hive version	Port number
Hive 0.13.1	10000

Hive version	Port number
Hive 0.11.0	10004
Hive 0.8.1	10003
Hive 0.7.1	10002
Hive 0.7	10001
Hive 0.5	10000

For more information about thrift services, see http://wiki.apache.org/thrift/.

Use Hive to recover partitions

Amazon EMR includes a statement in the Hive query language that recovers the partitions of a table from table data located in Amazon S3. The following example shows this.

```
CREATE EXTERNAL TABLE (json string) raw_impression
PARTITIONED BY (dt string)
LOCATION 's3://elastic-mapreduce/samples/hive-ads/tables/impressions';
ALTER TABLE logs RECOVER PARTITIONS;
```

The partition directories and data must be at the location specified in the table definition and must be named according to the Hive convention: for example, `dt=2009-01-01`.

> **Note**
>
> After Hive 0.13.1 this capability is supported natively using `msck repair` *`table`* and therefore `recover partitions` is not supported. For more information, see https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL.

Pass a Hive variable to a script

To pass a variable into a Hive step using the AWS CLI, type the following command, replace *mykey* with the name of your EC2 key pair, and replace *mybucket* with your bucket name. In this example, SAMPLE is a variable value preceded by the -d switch. This variable is defined in the Hive script as: ${SAMPLE}.

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.9 \
--applications Name=Hue Name=Hive Name=Pig \
--use-default-roles --ec2-attributes KeyName=myKey \
--instance-type m3.xlarge --instance-count 3 \
--steps Type=Hive,Name="Hive Program",ActionOnFailure=CONTINUE,\
Args=[-f,s3://elasticmapreduce/samples/hive-ads/libs/response-time-stats.q,-d,\
INPUT=s3://elasticmapreduce/samples/hive-ads/tables,-d,OUTPUT=s3://mybucket/hive-ads/
output/,\
-d,SAMPLE=s3://elasticmapreduce/samples/hive-ads/]
```

Specify an external metastore location

The following procedure shows you how to override the default configuration values for the Hive metastore location and start a cluster using the reconfigured metastore location.

To create a metastore located outside of the EMR cluster

1. Create a MySQL or Aurora database using Amazon RDS.

   For information about how to create an Amazon RDS database, see Getting started with Amazon RDS.

2. Modify your security groups to allow JDBC connections between your database and the **ElasticMapReduce-Master** security group.

   For information about how to modify your security groups for access, see Amazon RDS security groups in the *Amazon RDS User Guide*.

3. Set the JDBC configuration values in `hive-site.xml`:

   a. Create a `hive-site.xml` configuration file containing the following:

```
<configuration>
  <property>
    <name>javax.jdo.option.ConnectionURL</name>
    <value>jdbc:mariadb://hostname:3306/hive?createDatabaseIfNotExist=true</
value>
    <description>JDBC connect string for a JDBC metastore</description>
  </property>
  <property>
    <name>javax.jdo.option.ConnectionUserName</name>
    <value>hive</value>
    <description>Username to use against metastore database</description>
  </property>
  <property>
    <name>javax.jdo.option.ConnectionPassword</name>
    <value>password</value>
    <description>Password to use against metastore database</description>
  </property>
</configuration>
```

   `hostname` is the DNS address of the Amazon RDS instance running the database. *`username`* and *`password`* are the credentials for your database. For more information about connecting to MySQL and Aurora database instances, see [Connecting to a DB instance running the MySQL database engine] and [Connecting to an Aurora DB cluster] in the *Amazon RDS User Guide*.

   The JDBC drivers are installed by Amazon EMR.

   > **Note**
   >
   > The value property should not contain any spaces or carriage returns. It should appear all on one line.

   b. Save your `hive-site.xml` file to a location on Amazon S3, such as `s3://`*`mybucket/`*`hive-site.xml`.

4. Create a cluster, specifying the Amazon S3 location of the customized `hive-site.xml` file.

   The following example command demonstrates an AWS CLI command that does this.

Note

Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.10 \
--applications Name=Hue Name=Hive Name=Pig \
--use-default-roles --ec2-attributes KeyName=myKey \
--instance-type m3.xlarge --instance-count 3 \
--bootstrap-actions Name="Install Hive Site Configuration",\
Path="s3://region.elasticmapreduce/libs/hive/hive-script",\
Args=["--base-path","s3://elasticmapreduce/libs/hive","--install-hive-site",\
"--hive-site=s3://mybucket/hive-site.xml","--hive-versions","latest"]
```

Connect to Hive using JDBC

To connect to Hive via JDBC requires you to download the JDBC driver and install a SQL client. The following example demonstrates using SQL Workbench/J to connect to Hive using JDBC.

To download JDBC drivers

1. Download and extract the drivers appropriate to the versions of Hive that you want to access. The Hive version differs depending on the AMI that you choose when you create an Amazon EMR cluster.

   - Hive 0.13.1 JDBC drivers: https://amazon-odbc-jdbc-drivers.s3.amazonaws.com/public/AmazonHiveJDBC_1.0.4.1004.zip
   - Hive 0.11.0 JDBC drivers: https://mvnrepository.com/artifact/org.apache.hive/hive-jdbc/0.11.0
   - Hive 0.8.1 JDBC drivers: https://mvnrepository.com/artifact/org.apache.hive/hive-jdbc/0.8.1

2. Install SQL Workbench/J. For more information, see Installing and starting SQL Workbench/J in the SQL Workbench/J Manual User's Manual.

3. Create an SSH tunnel to the cluster master node. The port for connection is different depending on the version of Hive. Example commands are provided in the tables below for Linux `ssh` users and PuTTY commands for Windows users

Linux SSH commands

Hive version	Command
0.13.1	`ssh -o ServerAliveInterval=10 -i` *`path-to-key-file`* `-N -L 10000:localhost:10000 hadoop@` *`master-public-dns-name`*
0.11.0	`ssh -o ServerAliveInterval=10 -i` *`path-to-key-file`* `-N -L 10004:localhost:10004 hadoop@` *`master-public-dns-name`*
0.8.1	`ssh -o ServerAliveInterval=10 -i` *`path-to-key-file`* `-N -L 10003:localhost:10003 hadoop@` *`master-public-dns-name`*
0.7.1	`ssh -o ServerAliveInterval=10 -i` *`path-to-key-file`* `-N -L 10002:localhost:10002 hadoop@` *`master-public-dns-name`*
0.7	`ssh -o ServerAliveInterval=10 -i` *`path-to-key-file`* `-N -L 10001:localhost:10001 hadoop@` *`master-public-dns-name`*
0.5	`ssh -o ServerAliveInterval=10 -i` *`path-to-key-file`* `-N -L 10000:localhost:10000 hadoop@` *`master-public-dns-name`*

Windows PuTTY tunnel settings

Hive version	Tunnel settings
0.13.1	**Source port**: 10000 **Destination**: *`master-public-dns-name`* :10000
0.11.0	**Source port**: 10004 **Destination**: *`master-public-dns-name`* :10004
0.8.1	**Source port**: 10003 **Destination**: *`master-public-dns-name`* :10003

4. Add the JDBC driver to SQL Workbench.

   a. In the **Select Connection Profile** dialog box, choose **Manage Drivers**.

   b. Choose the **Create a new entry** (blank page) icon.

   c. In the **Name** field, type **`Hive JDBC`**.

d. For **Library**, click the **Select the JAR file(s)** icon.

e. Select JAR files as shown in the following table.

Hive driver version	JAR files to add

0.13.1	`hive_metastore.jar`
	`hive_service.jar`
	`HiveJDBC3.jar`
	`libfb303-0.9.0.jar`
	`libthrift-0.9.0.jar`
	`log4j-1.2.14.jar`
	`ql.jar`
	`slf4j-api-1.5.8.jar`
	`slf4j-log4j12-1.5.8.jar`
	`TCLIServiceClient.jar`

0.11.0	`hadoop-core-1.0.3.jar`
	`hive-exec-0.11.0.jar`
	`hive-jdbc-0.11.0.jar`
	`hive-metastore-0.11.0.jar`
	`hive-service-0.11.0.jar`
	`libfb303-0.9.0.jar`
	`commons-logging-1.0.4.jar`
	`slf4j-api-1.6.1.jar`

0.8.1	`hadoop-core-0.20.205.jar`
	`hive-exec-0.8.1.jar`
	`hive-jdbc-0.8.1.jar`
	`hive-metastore-0.8.1.jar`
	`hive-service-0.8.1.jar`
	`libfb303-0.7.0.jar`
	`libthrift-0.7.0.jar`
	`log4j-1.2.15.jar`
	`slf4j-api-1.6.1.jar`
	`slf4j-log4j12-1.6.1.jar`

0.7.1	`hadoop-0.20-core.jar`

Hive driver version	JAR files to add

	`hive-exec-0.7.1.jar`
	`hive-jdbc-0.7.1.jar`
	`hive-metastore-0.7.1.jar`
	`hive-service-0.7.1.jar`
	`libfb303.jar`
	`commons-logging-1.0.4.jar`
	`slf4j-api-1.6.1.jar`
	`slf4j-log4j12-1.6.1.jar`

0.7	`hadoop-0.20-core.jar`
	`hive-exec-0.7.0.jar`
	`hive-jdbc-0.7.0.jar`
	`hive-metastore-0.7.0.jar`
	`hive-service-0.7.0.jar`
	`libfb303.jar`
	`commons-logging-1.0.4.jar`
	`slf4j-api-1.5.6.jar`
	`slf4j-log4j12-1.5.6.jar`

0.5	`hadoop-0.20-core.jar`
	`hive-exec-0.5.0.jar`
	`hive-jdbc-0.5.0.jar`
	`hive-metastore-0.5.0.jar`
	`hive-service-0.5.0.jar`
	`libfb303.jar`
	`log4j-1.2.15.jar`
	`commons-logging-1.0.4.jar`

f. In the **Please select one driver** dialog box, select a driver according to the following table and click **OK**.

Hive version	Driver classname
0.13.1	`com.amazon.hive.jdbc3.HS2Driver`
0.11.0	`org.apache.hadoop.hive.jdbc.HiveDriver.jar`
0.8.1	`org.apache.hadoop.hive.jdbc.HiveDriver.jar`
0.7.1	`org.apache.hadoop.hive.jdbc.HiveDriver.jar`
0.7	`org.apache.hadoop.hive.jdbc.HiveDriver.jar`
0.5	`org.apache.hadoop.hive.jdbc.HiveDriver.jar`

5. When you return to the **Select Connection Profile** dialog box, verify that the **Driver** field is set to **Hive JDBC** and provide the JDBC connection string in the **URL** field according to the following table.

Hive version	JDBC connection string
0.13.1	`jdbc:hive2://localhost:10000/default`
0.11.0	`jdbc:hive://localhost:10004/default`
0.8.1	`jdbc:hive://localhost:10003/default`

If your cluster uses AMI version 3.3.1 or later, in the **Select Connection Profile** dialog box, type **hadoop** in the **Username** field.

HBase application specifics for earlier AMI versions of Amazon EMR

Supported HBase versions

HBase version	AMI version	AWS CLI configura tion parameters	HBase version details

0.94.18	3.1.0 and later	`--ami-version 3.1`
		`--ami-version 3.2`
		`--ami-version 3.3`
		`--applications Name=HBase`
			• Bug fixes and enhancements.

0.94.7	3.0-3.0.4	`--ami-version 3.0` `--applications Name=HBase`	
0.92	2.2 and later	`--ami-version 2.2 or later` `--applications Name=HBase`	

HBase cluster prerequisites

A cluster created using Amazon EMR AMI versions 2.x and 3.x should meet the following requirements for HBase.

- The AWS CLI (optional)—To interact with HBase using the command line, download and install the latest version of the AWS CLI. For more information, see Installing the AWS Command Line Interface in the *AWS Command Line Interface User Guide*.

- At least two instances (optional)—The cluster's master node runs the HBase master server and Zookeeper, and task nodes run the HBase region servers. For best performance, HBase clusters should run on at least two EC2 instances, but you can run HBase on a single node for evaluation purposes.
- Long-running cluster—HBase only runs on long-running clusters. By default, the CLI and Amazon EMR console create long-running clusters.
- An Amazon EC2 key pair set (recommended)—To use the Secure Shell (SSH) network protocol to connect with the master node and run HBase shell commands, you must use an Amazon EC2 key pair when you create the cluster.
- The correct AMI and Hadoop versions—HBase clusters are currently supported only on Hadoop 20.205 or later.
- Ganglia (optional)—To monitor HBase performance metrics, install Ganglia when you create the cluster.
- An Amazon S3 bucket for logs (optional)—The logs for HBase are available on the master node. If you'd like these logs copied to Amazon S3, specify an S3 bucket to receive log files when you create the cluster.

Creating a cluster with HBase

The following table lists options that are available when using the console to create a cluster with HBase using an Amazon EMR AMI release version.

Field	Action
Restore from backup	Specify whether to pre-load the HBase cluster with data stored in Amazon S3.
Backup location	Specify the URI where the backup from which to restore resides in Amazon S3.
Backup version	Optionally, specify the version name of the backup at **Backup Location** to use. If you leave this field blank, Amazon EMR uses the latest backup at **Backup Location** to populate the new HBase cluster.

Field	Action
Schedule Regular Backups	Specify whether to schedule automatic incremental backups. The first backup is a full backup to create a baseline for future incremental backups.
Consistent backup	Specify whether the backups should be consistent. A consistent backup is one that pauses write operations during the initial backup stage, synchroni zation across nodes. Any write operations thus paused are placed in a queue and resume when synchronization completes.
Backup frequency	The number of days/hours/minutes between scheduled backups.
Backup location	The Amazon S3 URI where backups are stored. The backup location for each HBase cluster should be different to ensure that differential backups stay correct.
Backup start time	Specify when the first backup should occur. You can set this to now, which causes the first backup to start as soon as the cluster is running, or enter a date and time in ISO format. For example, 2012-06-15T20:00Z would set the start time to June 15, 2012 at 8PM UTC.

The following example AWS CLI command launches a cluster with HBase and other applications:

Note

Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.3 \
               --applications Name=Hue Name=Hive Name=Pig Name=HBase \
               --use-default-roles --ec2-attributes KeyName=myKey \
               --instance-type c1.xlarge --instance-count 3 --termination-protected
```

After the connection between the Hive and HBase clusters has been made (as shown in the previous procedure), you can access the data stored on the HBase cluster by creating an external table in Hive.

The following example, when run from the Hive prompt, creates an external table that references data stored in an HBase table called `inputTable`. You can then reference `inputTable` in Hive statements to query and modify data stored in the HBase cluster.

> **Note**
>
> The following example uses **protobuf-java-2.4.0a.jar** in AMI 2.3.3, but you should modify the example to match your version. To check which version of the Protocol Buffers JAR you have, run the command at the Hive command prompt: `! ls /home/hadoop/lib;`.

```
add jar lib/emr-metrics-1.0.jar ;
               add jar lib/protobuf-java-2.4.0a.jar ;

               set hbase.zookeeper.quorum=ec2-107-21-163-157.compute-1.amazonaws.com ;

               create external table inputTable (key string, value string)
                    stored by 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
                     with serdeproperties ("hbase.columns.mapping" = ":key,f1:col1")
                     tblproperties ("hbase.table.name" = "t1");

               select count(*) from inputTable ;
```

Customizing HBase configuration

Although the default settings should work for most applications, you have the flexibility to modify your HBase configuration settings. To do this, run one of two bootstrap action scripts:

- **configure-hbase-daemons**—Configures properties of the master, regionserver, and zookeeper daemons. These properties include heap size and options to pass to the Java Virtual Machine (JVM) when the HBase daemon starts. You set these properties as arguments in the bootstrap action. This bootstrap action modifies the /home/hadoop/conf/hbase-user-env.sh configuration file on the HBase cluster.
- **configure-hbase**—Configures HBase site-specific settings such as the port the HBase master should bind to and the maximum number of times the client CLI client should retry an action.

You can set these one-by-one, as arguments in the bootstrap action, or you can specify the location of an XML configuration file in Amazon S3. This bootstrap action modifies the /home/hadoop/conf/hbase-site.xml configuration file on the HBase cluster.

> **Note**
>
> These scripts, like other bootstrap actions, can only be run when the cluster is created; you cannot use them to change the configuration of an HBase cluster that is currently running.

When you run the **configure-hbase** or **configure-hbase-daemons** bootstrap actions, the values you specify override the default values. Any values that you don't explicitly set receive the default values.

Configuring HBase with these bootstrap actions is analogous to using bootstrap actions in Amazon EMR to configure Hadoop settings and Hadoop daemon properties. The difference is that HBase does not have per-process memory options. Instead, memory options are set using the `--*daemon*-opts` argument, where `*daemon*` is replaced by the name of the daemon to configure.

Configure HBase daemons

Amazon EMR provides a bootstrap action, `s3://*region*.elasticmapreduce/bootstrap-actions/configure-hbase-daemons`, that you can use to change the configuration of HBase daemons, where *region* is the region into which you're launching your HBase cluster.

To configure HBase daemons using the AWS CLI, add the `configure-hbase-daemons` bootstrap action when you launch the cluster to configure one or more HBase daemons. You can set the following properties.

Property	Description
`hbase-master-opts`	Options that control how the JVM runs the master daemon. If set, these override the default HBASE_MASTER_OPTS variables.
`regionserver-opts`	Options that control how the JVM runs the region server daemon. If set, these override the default HBASE_REGIONSERVER_OPTS variables.

Property	Description
`zookeeper-opts`	Options that control how the JVM runs the zookeeper daemon. If set, these override the default HBASE_ZOOKEEPER_OPTS variables.

For more information about these options, see hbase-env.sh in the HBase documentation.

A bootstrap action to configure values for `zookeeper-opts` and `hbase-master-opts` is shown in the following example.

Note

Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.3 \
--applications Name=Hue Name=Hive Name=Pig Name=HBase \
--use-default-roles --ec2-attributes KeyName=myKey \
--instance-type c1.xlarge --instance-count 3 --termination-protected \
--bootstrap-actions Path=s3://elasticmapreduce/bootstrap-actions/configure-hbase-daemons,\
Args=["--hbase-zookeeper-opts=-Xmx1024m -XX:GCTimeRatio=19","--hbase-master-opts=-Xmx2048m","--hbase-regionserver-opts=-Xmx4096m"]
```

Configure HBase site settings

Amazon EMR provides a bootstrap action, `s3://elasticmapreduce/bootstrap-actions/configure-hbase`, that you can use to change the configuration of HBase. You can set configuration values one-by-one, as arguments in the bootstrap action, or you can specify the location of an XML configuration file in Amazon S3. Setting configuration values one-by-one is useful if you only need to set a few configuration settings. Setting them using an XML file is useful if you have many changes to make, or if you want to save your configuration settings for reuse.

Note

You can prefix the Amazon S3 bucket name with a region prefix, such as `s3://region.elasticmapreduce/bootstrap-actions/configure-hbase`, where *region* is the region into which you're launching your HBase cluster.

This bootstrap action modifies the `/home/hadoop/conf/hbase-site.xml` configuration file on the HBase cluster. The bootstrap action can only be run when the HBase cluster is launched.

For more information about the HBase site settings that you can configure, see Default configuration in the HBase documentation.

Set the `configure-hbase` bootstrap action when you launch the HBase cluster and specify the values in `hbase-site.xml` to change.

To specify individual HBase site settings using the AWS CLI

- To change the `hbase.hregion.max.filesize` setting, type the following command and replace *myKey* with the name of your Amazon EC2 key pair.

  Note

  Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.3 \
--applications Name=Hue Name=Hive Name=Pig Name=HBase \
--use-default-roles --ec2-attributes KeyName=myKey \
--instance-type c1.xlarge --instance-count 3 --termination-protected \
--bootstrap-actions Path=s3://elasticmapreduce/bootstrap-actions/configure-
hbase,Args=["-s","hbase.hregion.max.filesize=52428800"]
```

To specify HBase site settings with an XML file using the AWS CLI

1. Create a custom version of `hbase-site.xml`. Your custom file must be valid XML. To reduce the chance of introducing errors, start with the default copy of `hbase-site.xml`, located on

the Amazon EMR HBase master node at `/home/hadoop/conf/hbase-site.xml`, and edit a copy of that file instead of creating a file from scratch. You can give your new file a new name, or leave it as `hbase-site.xml`.

2. Upload your custom `hbase-site.xml` file to an Amazon S3 bucket. It should have permissions set so the AWS account that launches the cluster can access the file. If the AWS account launching the cluster also owns the Amazon S3 bucket, it has access.
3. Set the **configure-hbase** bootstrap action when you launch the HBase cluster, and include the location of your custom `hbase-site.xml` file. The following example sets the HBase site configuration values to those specified in the file `s3://mybucket/my-hbase-site.xml`. Type the following command, replace *myKey* with the name of your EC2 key pair, and replace *mybucket* with the name of your Amazon S3 bucket.

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.3 \
        --applications Name=Hue Name=Hive Name=Pig Name=HBase \
        --use-default-roles --ec2-attributes KeyName=myKey \
        --instance-type c1.xlarge --instance-count 3 --termination-protected \
        --bootstrap-actions Path=s3://elasticmapreduce/bootstrap-actions/configure-
hbase,Args=["--site-config-file","s3://mybucket/config.xml"]
```

If you specify more than one option to customize HBase operation, you must prepend each key-value pair with a `-s` option switch, as shown in the following example:

```
       --bootstrap-actions s3://elasticmapreduce/bootstrap-actions/configure-
hbase,Args=["-s","zookeeper.session.timeout=60000"]
```

With the proxy set and the SSH connection open, you can view the HBase UI by opening a browser window with http://*master-public-dns-name*:60010/master-status, where *master-public-dns-name* is the public DNS address of the master node in the HBase cluster.

You can view the current HBase logs by using SSH to connect to the master node, and navigating to the `mnt/var/log/hbase` directory. These logs are not available after the cluster is terminated unless you enable logging to Amazon S3 when the cluster is launched.

Back up and restore HBase

Amazon EMR provides the ability to back up your HBase data to Amazon S3, either manually or on an automated schedule. You can perform both full and incremental backups. After you have a backed-up version of HBase data, you can restore that version to an HBase cluster. You can restore to an HBase cluster that is currently running, or launch a new cluster pre-populated with backed-up data.

During the backup process, HBase continues to execute write commands. Although this ensures that your cluster remains available throughout the backup, there is the risk of inconsistency between the data being backed up and any write operations being executed in parallel. To understand the inconsistencies that might arise, you have to consider that HBase distributes write operations across the nodes in its cluster. If a write operation happens after a particular node is polled, that data is not included in the backup archive. You may even find that earlier writes to the HBase cluster (sent to a node that has already been polled) might not be in the backup archive, whereas later writes (sent to a node before it was polled) are included.

If a consistent backup is required, you must pause writes to HBase during the initial portion of the backup process, synchronization across nodes. You can do this by specifying the `--consistent` parameter when requesting a backup. With this parameter, writes during this period are queued and executed as soon as the synchronization completes. You can also schedule recurring backups, which resolves any inconsistencies over time, as data that is missed on one backup pass is backed up on the following pass.

When you back up HBase data, you should specify a different backup directory for each cluster. An easy way to do this is to use the cluster identifier as part of the path specified for the backup directory. For example, `s3://mybucket/backups/j-3AEXXXXXX16F2`. This ensures that any future incremental backups reference the correct HBase cluster.

When you are ready to delete old backup files that are no longer needed, we recommend that you first do a full backup of your HBase data. This ensures that all data is preserved and provides a baseline for future incremental backups. After the full backup is done, you can navigate to the backup location and manually delete the old backup files.

The HBase backup process uses S3DistCp for the copy operation, which has certain limitations regarding temporary file storage space.

Back up and restore HBase using the console

The console provides the ability to launch a new cluster and populate it with data from a previous HBase backup. It also gives you the ability to schedule periodic incremental backups of HBase data. Additional backup and restore functionality, such as the ability to restore data to an already running cluster, do manual backups, and schedule automated full backups, is available using the CLI.

To populate a new cluster with archived HBase data using the console

1. Navigate to the new Amazon EMR console and select **Switch to the old console** from the side navigation. For more information on what to expect when you switch to the old console, see Using the old console.
2. Choose **Create cluster**.
3. In the **Software Configuration** section, for **Additional Applications**, choose **HBase** and **Configure and add**.
4. On the **Add Application** dialog box, check **Restore From Backup**.
5. For **Backup Location**, specify the location of the backup yto load into the new HBase cluster. This should be an Amazon S3 URL of the form `s3://myawsbucket/backups/`.
6. For **Backup Version**, you have the option to specify the name of a backup version to load by setting a value. If you do not set a value for **Backup Version**, Amazon EMR loads the latest backup in the specified location.
7. Choose **Add** and proceed to create the cluster with other options as desired.

To schedule automated backups of HBase data using the console

1. In the **Software Configuration** section, for **Additional Applications**, choose **HBase** and **Configure and add**.
2. Choose **Schedule Regular Backups**.
3. Specify whether the backups should be consistent. A consistent backup is one that pauses write operations during the initial backup stage, synchronization across nodes. Any write operations thus paused are placed in a queue and resume when the synchronization completes.
4. Set how often backups should occur by entering a number for **Backup Frequency** and choosing **Days**, **Hours**, or **Minutes**. The first automated backup that runs is a full backup; after that, Amazon EMR saves incremental backups based on the schedule that you specify.

5. Specify the location in Amazon S3 where the backups should be stored. Each HBase cluster should be backed up to a separate location in Amazon S3 to ensure that incremental backups are calculated correctly.
6. Specify when the first backup should occur by setting a value for **Backup Start Time**. You can set this to `now`, which causes the first backup to start as soon as the cluster is running, or enter a date and time in ISO format. For example, 2013-09-26T20:00Z, sets the start time to September 26, 2013 at 8PM UTC.
7. Choose **Add**.
8. Proceed with creating the cluster with other options as desired.

Monitor HBase with CloudWatch

Amazon EMR reports three metrics to CloudWatch that you can use to monitor your HBase backups. These metrics are pushed to CloudWatch at five-minute intervals, and are provided without charge.

Metric	Description
`HBaseBackupFailed`	Whether the last backup failed. This is set to 0 by default and updated to 1 if the previous backup attempt failed. This metric is only reported for HBase clusters. Use case: Monitor HBase backups Units: *Count*
`HBaseMostRecentBackupDuration`	The amount of time it took the previous backup to complete. This metric is set regardless of whether the last completed backup succeeded or failed. While the backup is ongoing, this metric returns the number of minutes after the backup started. This metric is only reported for HBase clusters. Use case: Monitor HBase Backups Units: *Minutes*

Metric	Description
HBaseTimeSinceLast SuccessfulBackup	The number of elapsed minutes after the last successful HBase backup started on your cluster. This metric is only reported for HBase clusters. Use case: Monitor HBase backups Units: *Minutes*

Configure Ganglia for HBase

You configure Ganglia for HBase using the **configure-hbase-for-ganglia** bootstrap action. This bootstrap action configures HBase to publish metrics to Ganglia.

You must configure HBase and Ganglia when you launch the cluster; Ganglia reporting cannot be added to a running cluster.

Ganglia also stores log files on the server at `/mnt/var/log/ganglia/rrds`. If you configured your cluster to persist log files to an Amazon S3 bucket, the Ganglia log files are persisted there as well.

To launch a cluster with Ganglia for HBase, use the **configure-hbase-for-ganglia** bootstrap action as shown in the following example.

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.3 \
--applications Name=Hue Name=Hive Name=Pig Name=HBase Name=Ganglia \
--use-default-roles --ec2-attributes KeyName=myKey \
--instance-type c1.xlarge --instance-count 3 --termination-protected \
--bootstrap-actions Path=s3://elasticmapreduce/bootstrap-actions/configure-hbase-for-
ganglia
```

After the cluster is launched with Ganglia configured, you can access the Ganglia graphs and reports using the graphical interface running on the master node.

Pig application specifics for earlier AMI versions of Amazon EMR

Supported Pig versions

The Pig version you can add to your cluster depends on the version of the Amazon EMR AMI and the version of Hadoop you are using. The table below shows which AMI versions and versions of Hadoop are compatible with the different versions of Pig. We recommend using the latest available version of Pig to take advantage of performance enhancements and new functionality.

When you use the API to install Pig, the default version is used unless you specify `--pig-versions` as an argument to the step that loads Pig onto the cluster during the call to RunJobFlow.

Pig version	AMI version	Configuration parameters	Pig version details

0.12.0
 	
Release notes
 	
Documentation
	3.1.0 and later	`--ami-version 3.1`
		
		`--ami-version 3.2`
		
		`--ami-version 3.3`
			Adds support for the following:
			
			• Streaming UDFs without JVM implementations
			• ASSERT and IN operators
			• CASE expression
			• AvroStorage as a Pig built-in function.
			• ParquetLoader and ParquetStorer as built-in functions
			• BigInteger and BigDecimal types

Pig version	AMI version	Configuration parameters	Pig version details

0.11.1.1
 	
Release notes
 	
Documentation
	2.2 and later	`--pig-versions 0.11.1.1`
		
		`--ami-version 2.2`
			Improves performance of LOAD command with PigStorage if input resides in Amazon S3.

0.11.1
 	
Release notes
 	
Documentation
	2.2 and later	`--pig-versions 0.11.1`
		
		`--ami-version 2.2`
			Adds support for JDK 7, Hadoop 2, Groovy user-defined functions, SchemaTuple optimization, new operators, and more. For more information, see Pig 0.11.1 change log.

0.9.2.2
 	
Release notes
 	
Documentation
	2.2 and later	`--pig-versions 0.9.2.2`
		
		`--ami-version 2.2`
			Adds support for Hadoop 1.0.3.

0.9.2.1
 	
Release notes
 	
Documentation
	2.2 and later	`--pig-versions 0.9.2.1`
		
		`--ami-version 2.2`
			Adds support for MapR.

Pig version	AMI version	Configuration parameters	Pig version details
0.9.2 Release notes Documentation	2.2 and later	`--pig-versions 0.9.2` `--ami-version 2.2`	Includes several performance improvements and bug fixes. For complete information about the changes for Pig 0.9.2, go to the Pig 0.9.2 change log.
0.9.1 Release notes Documentation	2.0	`--pig-versions 0.9.1` `--ami-version 2.0`	
0.6 Release notes	1.0	`--pig-versions 0.6` `--ami-version 1.0`	
0.3 Release notes	1.0	`--pig-versions 0.3` `--ami-version 1.0`	

Pig version details

Amazon EMR supports certain Pig releases that might have additional Amazon EMR patches applied. You can configure which version of Pig to run on Amazon EMR clusters. For more information about how to do this, see Apache Pig. The following sections describe different Pig versions and the patches applied to the versions loaded on Amazon EMR.

Pig patches

This section describes the custom patches applied to Pig versions available with Amazon EMR.

Pig 0.11.1.1 patches

The Amazon EMR version of Pig 0.11.1.1 is a maintenance release that improves performance of LOAD command with PigStorage if the input resides in Amazon S3.

Pig 0.11.1 patches

The Amazon EMR version of Pig 0.11.1 contains all the updates provided by the Apache Software Foundation and the cumulative Amazon EMR patches from Pig version 0.9.2.2. However, there are no new Amazon EMR-specific patches in Pig 0.11.1.

Pig 0.9.2 patches

Apache Pig 0.9.2 is a maintenance release of Pig. The Amazon EMR team has applied the following patches to the Amazon EMR version of Pig 0.9.2.

Patch	Description
PIG-1429	Add the Boolean data type to Pig as a first class data type. For more information, go to https://issues.apache.org/jira/browse/PIG-1429. **Status:** Committed **Fixed in Apache Pig Version:** 0.10
PIG-1824	Support import modules in Jython UDF. For more information, go to https://issues.apache.org/jira/browse/PIG-1824. **Status:** Committed **Fixed in Apache Pig Version:** 0.10
PIG-2010	Bundle registered JARs on the distributed cache. For more information, go to https://issues.apache.org/jira/browse/PIG-2010. **Status:** Committed

Patch	Description
	Fixed in Apache Pig Version: 0.11
PIG-2456	Add a ~/.pigbootup file where the user can specify default Pig statements. For more information, go to https://issues.apache.org/jira/browse/PIG-2456. **Status:** Committed **Fixed in Apache Pig Version:** 0.11
PIG-2623	Support using Amazon S3 paths to register UDFs. For more information, go to https://issues.apache.org/jira/browse/PIG-2623. **Status:** Committed **Fixed in Apache Pig Version:** 0.10, 0.11

Pig 0.9.1 patches

The Amazon EMR team has applied the following patches to the Amazon EMR version of Pig 0.9.1.

Patch	Description
Support JAR files and Pig scripts in dfs	Add support for running scripts and registering JAR files stored in HDFS, Amazon S3, or other distributed file systems. For more information, go to https://issues.apache.org/jira/browse/PIG-1505. **Status:** Committed **Fixed in Apache Pig Version:** 0.8.0
Support multiple file systems in Pig	Add support for Pig scripts to read data from one file system and write it to another. For more information, go to https://issues.apache.org/jira/browse/PIG-1564. **Status:** Not Committed

Patch	Description
	Fixed in Apache Pig Version: n/a
Add Piggybank datetime and string UDFs	Add datetime and string UDFs to support custom Pig scripts. For more information, go to https://issues.apache.org/jira/browse/PIG-1565. **Status:** Not Committed **Fixed in Apache Pig Version:** n/a

Interactive and batch Pig clusters

Amazon EMR enables you to run Pig scripts in two modes:

- Interactive
- Batch

When you launch a long-running cluster using the console or the AWS CLI, you can connect using **ssh** into the master node as the Hadoop user and use the Grunt shell to develop and run your Pig scripts interactively. Using Pig interactively enables you to revise the Pig script more easily than batch mode. After you successfully revise the Pig script in interactive mode, you can upload the script to Amazon S3 and use batch mode to run the script in production. You can also submit Pig commands interactively on a running cluster to analyze and transform data as needed.

In batch mode, you upload your Pig script to Amazon S3, and then submit the work to the cluster as a step. Pig steps can be submitted to a long-running cluster or a transient cluster.

Spark application specifics with earlier AMI versions of Amazon EMR

Use Spark interactively or in batch mode

Amazon EMR enables you to run Spark applications in two modes:

- Interactive
- Batch

When you launch a long-running cluster using the console or the AWS CLI, you can connect using SSH into the master node as the Hadoop user and use the Spark shell to develop and run your Spark applications interactively. Using Spark interactively enables you to prototype or test Spark applications more easily than in a batch environment. After you successfully revise the Spark application in interactive mode, you can put that application JAR or Python program in the file system local to the master node of the cluster on Amazon S3. You can then submit the application as a batch workflow.

In batch mode, upload your Spark script to Amazon S3 or the local master node file system, and then submit the work to the cluster as a step. Spark steps can be submitted to a long-running cluster or a transient cluster.

Creating a cluster with Spark installed

To launch a cluster with Spark installed using the console

1. Navigate to the new Amazon EMR console and select **Switch to the old console** from the side navigation. For more information on what to expect when you switch to the old console, see Using the old console.
2. Choose **Create cluster**.
3. For **Software Configuration**, choose the AMI release version that you require.
4. For **Applications to be installed**, choose **Spark** from the list, then choose **Configure and add**.
5. Add arguments to change the Spark configuration as desired. For more information, see Configure Spark. Choose **Add**.
6. Select other options as necessary and then choose **Create cluster**.

The following example shows how to create a cluster with Spark using Java:

```
AmazonElasticMapReduceClient emr = new AmazonElasticMapReduceClient(credentials);
SupportedProductConfig sparkConfig = new SupportedProductConfig()
   .withName("Spark");

RunJobFlowRequest request = new RunJobFlowRequest()
   .withName("Spark Cluster")
   .withAmiVersion("3.11.0")
   .withNewSupportedProducts(sparkConfig)
   .withInstances(new JobFlowInstancesConfig()
    .withEc2KeyName("myKeyName")
```

```
     .withInstanceCount(1)
     .withKeepJobFlowAliveWhenNoSteps(true)
     .withMasterInstanceType("m3.xlarge")
     .withSlaveInstanceType("m3.xlarge")
    );
RunJobFlowResult result = emr.runJobFlow(request);
```

Configure Spark

You configure Spark when you create a cluster by running the bootstrap action located at awslabs/emr-bootstrap-actions/spark repository on Github. For arguments that the bootstrap action accepts, see the READMEin that repository. The bootstrap action configures properties in the `$SPARK_CONF_DIR/spark-defaults.conf` file. For more information about settings, see the Spark Configuration topic in Spark documentation. You can replace "latest" in the following URL with the version number of Spark that you are installing, for example, `2.2.0` http://spark.apache.org/docs/latest/configuration.html.

You can also configure Spark dynamically at the time of each application submission. A setting to automatically maximize the resource allocation for an executor is available using the `spark` configuration file. For more information, see Overriding Spark default configuration settings.

Changing Spark default settings

The following example shows how to create a cluster with `spark.executor.memory` set to 2G using the AWS CLI.

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --name "Spark cluster" --ami-version 3.11.0 \
--applications Name=Spark, Args=[-d,spark.executor.memory=2G] --ec2-attributes
 KeyName=myKey \
--instance-type m3.xlarge --instance-count 3 --use-default-roles
```

Submit work to Spark

To submit work to a cluster, use a step to run the `spark-submit` script on your EMR cluster. Add the step using the `addJobFlowSteps` method in [AmazonElasticMapReduceClient]():

```
AWSCredentials credentials = new BasicAWSCredentials(accessKey, secretKey);
AmazonElasticMapReduceClient emr = new AmazonElasticMapReduceClient(credentials);
StepFactory stepFactory = new StepFactory();
AddJobFlowStepsRequest req = new AddJobFlowStepsRequest();
req.withJobFlowId("j-1K48XXXXXXHCB");

List<StepConfig> stepConfigs = new ArrayList<StepConfig>();

StepConfig sparkStep = new StepConfig()
 .withName("Spark Step")
 .withActionOnFailure("CONTINUE")
 .withHadoopJarStep(stepFactory.newScriptRunnerStep("/home/hadoop/spark/bin/spark-
submit","--class","org.apache.spark.examples.SparkPi","/home/hadoop/spark/lib/spark-
examples-1.3.1-hadoop2.4.0.jar","10"));

stepConfigs.add(sparkStep);
req.withSteps(stepConfigs);
AddJobFlowStepsResult result = emr.addJobFlowSteps(req);
```

Overriding Spark default configuration settings

You may want to override Spark default configuration values on a per-application basis. You can do this when you submit applications using a step, which essentially passes options to `spark-submit`. For example, you may wish to change the memory allocated to an executor process by changing `spark.executor.memory`. You can supply the `--executor-memory` switch with an argument like the following:

```
/home/hadoop/spark/bin/spark-submit --executor-memory 1g --class
 org.apache.spark.examples.SparkPi /home/hadoop/spark/lib/spark-examples*.jar 10
```

Similarly, you can tune `--executor-cores` and `--driver-memory`. In a step, you would provide the following arguments to the step:

```
--executor-memory 1g --class org.apache.spark.examples.SparkPi /home/hadoop/spark/lib/
spark-examples*.jar 10
```

You can also tune settings that may not have a built-in switch using the `--conf` option. For more information about other settings that are tunable, see the [Dynamically loading Spark properties](#) topic in the Apache Spark documentation.

S3DistCp utility differences with earlier AMI versions of Amazon EMR

S3DistCp versions supported in Amazon EMR

The following S3DistCp versions are supported in Amazon EMR AMI releases. S3DistCp versions after 1.0.7 are found on directly on the clusters. Use the JAR in `/home/hadoop/lib` for the latest features.

Version	Description	Release date
1.0.8	Adds the `--appendToLastFile` , `--requirePreviousManifest` , and `--storageClass` options.	3 January 2014
1.0.7	Adds the `--s3ServerSideEncryption` option.	2 May 2013
1.0.6	Adds the `--s3Endpoint` option.	6 August 2012
1.0.5	Improves the ability to specify which version of S3DistCp to run.	27 June 2012
1.0.4	Improves the `--deleteOnSuccess` option.	19 June 2012
1.0.3	Adds support for the `--numberFiles` and `--startingIndex` options.	12 June 2012
1.0.2	Improves file naming when using groups.	6 June 2012
1.0.1	Initial release of S3DistCp.	19 January 2012

Add an S3DistCp copy step to a cluster

To add an S3DistCp copy step to a running cluster, type the following command, replace *j-3GYXXXXXX9IOK* with your cluster ID, and replace *mybucket* with your Amazon S3 bucket name.

Note

Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr add-steps --cluster-id j-3GYXXXXXX9IOK \
--steps Type=CUSTOM_JAR,Name="S3DistCp step",Jar=/home/hadoop/lib/emr-s3distcp-1.0.jar,
\
Args=["--s3Endpoint,s3-eu-west-1.amazonaws.com",\
"--src,s3://mybucket/logs/j-3GYXXXXXX9IOJ/node/",\
"--dest,hdfs:///output",\
"--srcPattern,.*[a-zA-Z,]+"]
```

Example Load Amazon CloudFront logs into HDFS

This example loads Amazon CloudFront logs into HDFS by adding a step to a running cluster. In the process, it changes the compression format from Gzip (the CloudFront default) to LZO. This is useful because data compressed using LZO can be split into multiple maps as it is decompressed, so you don't have to wait until the compression is complete, as you do with Gzip. This provides better performance when you analyze the data using Amazon EMR. This example also improves performance by using the regular expression specified in the `--groupBy` option to combine all of the logs for a given hour into a single file. Amazon EMR clusters are more efficient when processing a few, large, LZO-compressed files than when processing many, small, Gzip-compressed files. To split LZO files, you must index them and use the hadoop-lzo third-party library.

To load Amazon CloudFront logs into HDFS, type the following command, replace *j-3GYXXXXXX9IOK* with your cluster ID, and replace *mybucket* with your Amazon S3 bucket name.

Note

Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr add-steps --cluster-id j-3GYXXXXXX9IOK \
--steps Type=CUSTOM_JAR,Name="S3DistCp step",Jar=/home/hadoop/lib/emr-s3distcp-1.0.jar,
\
```

```
Args=["--src,s3://mybucket/cf","--dest,hdfs:///local",\
"--groupBy,.*XABCD12345678.([0-9]+-[0-9]+-[0-9]+-[0-9]+).*",\
"--targetSize,128",
"--outputCodec,lzo","--deleteOnSuccess"]
```

Consider the case in which the preceding example is run over the following CloudFront log files.

```
s3://amzn-s3-demo-bucket1/cf/XABCD12345678.2012-02-23-01.HLUS3JKx.gz
s3://amzn-s3-demo-bucket1/cf/XABCD12345678.2012-02-23-01.I9CNAZrg.gz
s3://amzn-s3-demo-bucket1/cf/XABCD12345678.2012-02-23-02.YRRwERSA.gz
s3://amzn-s3-demo-bucket1/cf/XABCD12345678.2012-02-23-02.dshVLXFE.gz
s3://amzn-s3-demo-bucket1/cf/XABCD12345678.2012-02-23-02.LpLfuShd.gz
```

S3DistCp copies, concatenates, and compresses the files into the following two files, where the file name is determined by the match made by the regular expression.

```
hdfs:///local/2012-02-23-01.lzo
hdfs:///local/2012-02-23-02.lzo
```

What's new?

This page describes the changes and functionality available in the latest releases of Amazon EMR 7.x, 6.x, and 5.x.

These release notes are also available on the Amazon EMR 7.2.0, Amazon EMR 6.15.0, and Amazon EMR 5.36.2 pages, along with the application versions, component versions, and available configuration classifications for each release.

- For release notes from prior releases, see the Amazon EMR archive of release notes.
- To get updates when a new Amazon EMR release is available, subscribe to the RSS feed for Amazon EMR release notes.

> **Note**
>
> Later releases of Amazon EMR use AWS Signature Version 4 (SigV4) to authenticate requests to Amazon S3. We recommend that you use an Amazon EMR release that supports SigV4 so that you can access new S3 buckets and avoid interruption to your workloads. For more information and a list of Amazon EMR releases that support SigV4, see Amazon EMR and AWS Signature Version 4.

Amazon EMR 7.2.0 (latest release of 7.x series)

New Amazon EMR releases are made available in different Regions over a period of several days, beginning with the first Region on the initial release date. The latest release version may not be available in your Region during this period.

The following release notes include information for Amazon EMR release 7.2.0. Changes are relative to 7.2.0.

New features

- **Application upgrades** – Amazon EMR 7.2.0 application upgrades include Iceberg 1.5.0-amzn-0 and Delta 3.1.0.
- Amazon EMR adds support so that you can use other applications such as HBase, Flink, and Hive with the Amazon S3 Express One Zone storage class.

- This release adds the capability to read restored objects, so you can read Glacier objects from an S3 location with the S3A protocol. This capability works with Spark, Flink, and Hive.
- Node labels with managed scaling - Use node labels with your instances based on market type or node type to improve automatic scaling from Amazon EMR. For more information, see Using managed scaling in Amazon EMR.

Known issues

- Python 3.11 isn't supported with EMR Studio.

Changes, enhancements, and resolved issues

- This release fixes a deadlock issue that can occur during internal step cleanup operations. This operation manages the life cycle of steps as they complete on the EMR cluster. This issue affects critical Amazon EMR operations, such as step operation and scaling.
- This release resolves an issue where custom clusters with custom AMIs that have certain pre-existing log files can cause the Amazon EMR log management daemon to fail.
- Amazon EMR 7.2.0 upgrades the Amazon EMR daemon responsible for cluster management and monitoring activities from AWS SDK v1 to v2.

- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

OsRelea Label (Amazoi Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2023.5.2 240819.	6.1.102-111.182.am zn2023	August 20, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
			(Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2023.5.20240730.	6.1.97-104.177.amzn2023	August 2, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2023.5.20240722.	6.1.97-104.177.amzn2023	July 24, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2023.5.2 240708.	6.1.96-102.177.amzn2023	July 8th, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), Canada West (Calgary), AWS GovCloud (US-West), AWS GovCloud (US-East),China (Beijing), China (Ningxia)

Amazon EMR 6.15.0 (latest release of 6.x series)

New Amazon EMR releases are made available in different Regions over a period of several days, beginning with the first Region on the initial release date. The latest release version may not be available in your Region during this period.

The following release notes include information for Amazon EMR release 6.15.0. Changes are relative to 6.14.0. For information on the release timeline, see the 6.15.0 change log.

New features

- **Application upgrades** – Amazon EMR 6.15.0 application upgrades include Apache Hadoop 3.3.6, Apache Hudi 0.14.0-amzn-0, Iceberg 1.4.0-amzn-0, and Trino 426.
- **Faster launches for EMR clusters that run on EC2** – It's now up to 35% faster to launch an Amazon EMR on EC2 cluster. With this improvement, most customers can launch their clusters in 5 minutes or less.
- **CodeWhisperer for EMR Studio** – You can now use Amazon CodeWhisperer with Amazon EMR Studio to get real-time recommendations as you write code in JupyterLab. CodeWhisperer can complete your comments, finish single lines of code, make line-by-line recommendations, and generate fully-formed functions.
- **Faster job restart times with Flink** – With Amazon EMR 6.15.0 and higher, several new mechanisms are available for Apache Flink to improve the job restart time during task recovery or scaling operations. This optimizes the speed of recovery and restart of execution graphs to improve job stability.
- **Table-level and fine-grained access control for open-table formats** – With Amazon EMR 6.15.0 and higher, when you run Spark jobs on Amazon EMR on EC2 clusters that access data in the AWS Glue Data Catalog, you can use AWS Lake Formation to apply table, row, column, and cell level permissions on Hudi, Iceberg, or Delta Lake based tables.
- **Hadoop upgrade** – Amazon EMR 6.15.0 includes an upgrade of Apache Hadoop to version 3.3.6. Hadoop 3.3.6 was the latest version at the time of the Amazon EMR 6.15 deployment, released by Apache in June 2023. Prior releases of Amazon EMR (6.9.0 to 6.14.x) used Hadoop 3.3.3.

  The upgrade includes hundreds of improvements and fixes, and features that include reconfigurable datanode parameters, `DFSAdmin` option to initiate bulk reconfiguration operations on all live datanodes, and a vectored API that allows seek-heavy readers to specify multiple ranges to read. Hadoop 3.3.6 also adds support for HDFS APIs and semantics for its write-ahead log (WAL), so that HBase can run on other storage system implementations. For more information, see the changelogs for versions 3.3.4, 3.3.5, and 3.3.6 in the *Apache Hadoop documentation*.

- **Support for AWS SDK for Java, version 2** - Amazon EMR 6.15.0 applications can use AWS SDK for Java versions 1.12.569 or 2.20.160 if the application supports v2. The AWS SDK for Java 2.x is a major rewrite of the version 1.x code base. It's built on top of Java 8+ and adds several frequently requested features. These include support for non-blocking I/O, and the ability to plug in a different HTTP implementation at runtime. For more information, including a **Migration Guide from SDK for Java v1 to v2**, see the AWS SDK for Java, version 2 guide.

Changes, enhancements, and resolved issues

- To improve your high-availability EMR clusters, this release enables connectivity to Amazon EMR daemons on local host that use IPv6 endpoints.
- This release enables TLS 1.2 for communication with ZooKeeper provisioned on all the primary nodes of your high-availability cluster.
- This release improves the management of ZooKeeper transaction log files that are maintained on primary nodes to minimize scenarios where the log files grow out of bounds and interrupt cluster operations.
- This release makes intra-node communication more resilient for high-availability EMR clusters. This improvement reduces the chance of bootstrap action failures or cluster start failures.
- Tez in Amazon EMR 6.15.0 introduces configurations that you can specify to asynchronously open the input splits in a Tez grouped split. This results in faster performance of read queries when there are a large number of input splits in a single Tez grouped split. For more information, see Tez asynchronous split opening.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 816.0	4.14.350	August 21, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240809.0	4.14.349	August 20, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240719.0	4.14.348	July 25, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 709.1	4.14.348	July 23, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing) , China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240223.0	4.14.336	March 8, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 131.0	4.14.336	February 14, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240124.0	4.14.336	February 7, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240109.0	4.14.334	January 24, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231218.0	4.14.330	January 2, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231206.0	4.14.330	December 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231116.0	4.14.328	December 11, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231101.0	4.14.327	November 13, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

Amazon EMR 5.36.2 (latest release of 5.x series)

New Amazon EMR releases are made available in different Regions over a period of several days, beginning with the first Region on the initial release date. The latest release version may not be available in your Region during this period.

The following release notes include information for Amazon EMR release 5.36.2. Changes are relative to 5.36.1. For information on the release timeline, see the change log.

Changes, enhancements, and resolved issues

- This releases improves cluster scale-down logic so that Amazon EMR doesn't scale-down core nodes below the HDFS replication factor setting for the cluster. This improvement fulfills data redundancy requirements, and reduces the chance that a scaling operation might stall.
- This release adds a new retry mechanism to the cluster scaling workflow for that run Presto or Trino. This improvement reduces the risk that cluster resize runs indefinitely due to a single failed resize operation. It also improves cluster utilization, because your cluster scales up and down faster.
- Fixes an issue where cluster scale-down operations might stall while Amazon EMR gracefully decommissions a core node and it turns unhealthy before it is fully decommissioned.
- Improves the stability of a node in a high-availability cluster with multiple primary nodes when Amazon EMR restarts a single node.
- Optimizes log management with Amazon EMR running on Amazon EC2. As a result, you might see a slight reduction in storage costs for your cluster logs.
- Improves the management of ZooKeeper transaction log files that are maintained on primary nodes to minimize scenarios where the log files grow out of bounds and interrupt cluster operations.
- Fixes a rare bug which can cause a high-availability cluster with multiple primary nodes to fail due to not being able to communicate with the Yarn ResourceManager.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 816.0	4.14.350	August 21, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelease Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 809.0	4.14.349	August 20, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 719.0	4.14.348	July 25, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 503.0	4.14.343	xxxxxx, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

Amazon EMR and AWS Signature Version 4

Amazon EMR releases use AWS Signature Version 4 (SigV4) to authenticate requests to Amazon S3. Buckets created in Amazon S3 after June 24, 2020 don't support requests signed by Signature Version 2 (SigV2). Buckets created on or before June 24, 2020 will continue to support SigV2.

We recommend that you migrate to an Amazon EMR release that supports SigV4 so that you can access new S3 buckets and avoid interruption to your workloads.

If you use applications that are included with Amazon EMR such as Apache Spark, Apache Hive, and Presto, you don't need to change your application code to use SigV4 . If you use custom applications that are not included with Amazon EMR, you might need to update your code to use SigV4. For more information, see Moving from Signature Version 2 to Signature Version 4 in the Amazon S3 User Guide.

The following Amazon EMR releases support SigV4: emr-4.7.4, emr-4.8.5, emr-4.9.6, emr-4.10.1, emr-5.1.1, emr-5.2.3, emr-5.3.2, emr-5.4.1, emr-5.5.4, emr-5.6.1, emr-5.7.1, emr-5.8.3, emr-5.9.1, emr-5.10.1, emr-5.11.4, emr-5.12.3, emr-5.13.1, emr-5.14.2, emr-5.15.1, emr-5.16.1, emr-5.17.2, emr-5.18.1, emr-5.19.1, emr-5.20.1, emr-5.21.2, and emr-5.22.0 and higher. All 6.x and 7.x releases support SigV4.

Approach to mitigate CVE-2021-44228

Note

For Amazon EMR release 6.9.0 and later, all components installed by Amazon EMR that use Log4j libraries use Log4j version 2.17.1 or later.

Amazon EMR running on EC2

The issue discussed in CVE-2021-44228 is relevant to Apache Log4j core versions between 2.0.0 and 2.14.1 when processing inputs from untrusted sources. Amazon EMR clusters launched with Amazon EMR 5.x releases up to 5.34.0 and EMR 6.x releases up to Amazon EMR 6.5.0 include open-source frameworks such as Apache Hive, Flink, HUDI, Presto, and Trino, which use these versions of Apache Log4j. However, many customers use the open-source frameworks installed on their Amazon EMR clusters to process and log inputs from untrusted sources.

We recommend that you apply the "Amazon EMR Bootstrap Action Solution for Log4j CVE-2021-44228" as described in the following section. This solution also addresses CVE-2021-45046.

> **Note**
>
> The bootstrap action scripts for Amazon EMR were updated on September 7, 2022 to include incremental bug fixes and improvements for Oozie. If you use Oozie, you should apply the updated Amazon EMR bootstrap action solution described in the following section.

Amazon EMR on EKS

If you use Amazon EMR on EKS with default configuration, you are not impacted by the issue described in CVE-2021-44228, and you do not have to apply the solution described in the Amazon EMR bootstrap action solution for Log4j CVE-2021-44228 & CVE-2021-45046section. For Amazon EMR on EKS, the Amazon EMR runtime for Spark uses Apache Log4j version 1.2.17. When using Amazon EMR on EKS you should not change the default setting for `log4j.appender` component to `log`.

Amazon EMR bootstrap action solution for Log4j CVE-2021-44228 & CVE-2021-45046

This solution provides an Amazon EMR bootstrap action that must be applied on your Amazon EMR clusters. For each Amazon EMR release, you will find a link to a bootstrap action script below. To apply this bootstrap action, you should complete the following steps:

1. Copy the script that corresponds to your Amazon EMR release to a local S3 bucket in your AWS account. Please make sure that you are using a bootstrap script that is specific to your Amazon EMR release.
2. Set up a bootstrap action for your EMR clusters to run the script copied to your S3 bucket as per instructions described in EMR documentation. If you have other bootstrap actions configured for your EMR clusters, please ensure that this script is set up as the first bootstrap action script to execute.
3. Terminate existing EMR clusters, and launch new clusters with the bootstrap action script. AWS recommends that you test the bootstrap scripts in your test environment and validate your applications before applying it to your production environment. If you are not using the latest revision for an EMR minor release (for example, 6.3.0), you must use the latest revision (for example, 6.3.1), and then apply the solution discussed above.

CVE-2021-44228 & CVE-2021-45046 - Bootstrap Scripts for Amazon EMR Releases

Amazon EMR release number	Script location	Script release date
6.5.0	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-6.5.0-v2.sh`	March 24, 2022
6.4.0	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-6.4.0-v2.sh`	March 24, 2022
6.3.1	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-6.3.1-v2.sh`	March 24, 2022
6.2.1	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-6.2.1-v2.sh`	March 24, 2022
6.1.1	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-6.1.1-v2.sh`	December 14, 2021
6.0.1	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-6.0.1-v2.sh`	December 14, 2021
5.34.0	`s3://elasticmapreduce/bootstrap-actions/`	December 12, 2021

Amazon EMR release number	Script location	Script release date
	`log4j/patch-log4j-emr-5.34.0-v2.sh`	
5.33.1	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.33.1-v2.sh`	December 12, 2021
5.32.1	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.32.1-v2.sh`	December 13, 2021
5.31.1	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.31.1-v2.sh`	December 13, 2021
5.30.2	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.30.2-v2.sh`	December 14, 2021
5.29.0	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.29.0-v2.sh`	December 14, 2021
5.28.1	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.28.1-v2.sh`	December 15, 2021

Amazon EMR release number	Script location	Script release date
5.27.1	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.27.1-v2.sh`	December 15, 2021
5.26.0	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.26.0-v2.sh`	December 15, 2021
5.25.0	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.25.0-v2.sh`	December 15, 2021
5.24.1	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.24.1-v2.sh`	December 15, 2021
5.23.1	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.23.1-v2.sh`	December 15, 2021
5.22.0	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.22.0-v2.sh`	December 15, 2021
5.21.2	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.21.2-v2.sh`	December 15, 2021

Amazon EMR release number	Script location	Script release date
5.20.1	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.20.1-v2.sh`	December 15, 2021
5.19.1	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.19.1-v2.sh`	December 15, 2021
5.18.1	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.18.1-v2.sh`	December 15, 2021
5.17.2	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.17.2-v2.sh`	December 15, 2021
5.16.1	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.16.1-v2.sh`	December 15, 2021
5.15.1	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.15.1-v2.sh`	December 15, 2021
5.14.2	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.14.2-v2.sh`	December 15, 2021

Amazon EMR release number	Script location	Script release date
5.13.1	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.13.1-v2.sh`	December 15, 2021
5.12.3	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.12.3-v2.sh`	December 15, 2021
5.11.4	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.11.4-v2.sh`	December 15, 2021
5.10.1	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.10.1-v2.sh`	December 15, 2021
5.9.1	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.9.1-v2.sh`	December 15, 2021
5.8.3	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.8.3-v2.sh`	December 15, 2021
5.7.1	`s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.7.1-v2.sh`	December 15, 2021

EMR release version	Latest revision as of December 2021
6.3.0	6.3.1
6.2.0	6.2.1
6.1.0	6.1.1
6.0.0	6.0.1
5.33.0	5.33.1
5.32.0	5.32.1
5.31.0	5.31.1
5.30.0 or 5.30.1	5.30.2
5.28.0	5.28.1
5.27.0	5.27.1
5.24.0	5.24.1
5.23.0	5.23.1
5.21.0 or 5.21.1	5.21.2
5.20.0	5.20.1
5.19.0	5.19.1
5.18.0	5.18.1
5.17.0 or 5.17.1	5.17.2
5.16.0	5.16.1
5.15.0	5.15.1
5.14.0 or 5.14.1	5.14.2

EMR release version	Latest revision as of December 2021
5.13.0	5.13.1
5.12.0, 5.12.1, 5.12.2	5.12.3
5.11.0, 5.11.1, 5.11.2, 5.11.3	5.11.4
5.9.0	5.9.1
5.8.0, 5.8.1, 5.8.2	5.8.3
5.7.0	5.7.1

Frequently asked questions

- **Are EMR releases older than EMR 5 impacted by CVE-2021-44228**?

  No. EMR releases prior to EMR release 5 use Log4j versions older than 2.0.
- **Does this solution address CVE-2021-45046?**

  Yes, this solution also addresses **CVE-2021-45046.**
- **Does the solution handle custom applications that I install on my EMR clusters?**

  The bootstrap script only updates JAR files that are installed by EMR. If you install and run custom applications and JAR files on your EMR clusters through bootstrap actions, as steps submitted to your clusters, by using custom Amazon Linux AMI, or through any other mechanism, please work with your application vendor to determine if your custom applications are impacted by CVE-2021- 44228, and determine an appropriate solution.
- **How should I handle customized docker images with EMR on EKS?**

  If you add custom applications to Amazon EMR on EKS using customized docker images or submit jobs to Amazon EMR on EKSwith custom application files, please work with the application vendor to determine if your custom applications are impacted by CVE-2021-44228, and determine an appropriate solution.
- **How does the bootstrap script work to mitigate the issue described in CVE-2021-44228 and CVE-2021-45046?**

The bootstrap script updates EMR startup instructions by adding a new set of instructions. These new instructions delete the JndiLookup class files used through Log4j by all open source frameworks installed by EMR. This follows the recommendation published by Apache for addressing the Log4j issues.

- **Is there an update to EMR that uses Log4j versions 2.17.1 or higher?**

  EMR 5 releases up to release 5.34 and EMR 6 releases up to release 6.5 use older versions of open source frameworks that are incompatible with the latest versions of Log4j. If you continue to use these releases, we recommend that you apply the bootstrap action to mitigate the issues discussed in the CVEs. After EMR 5 release 5.34 and EMR 6 release 6.5, applications that use Log4j 1.x and Log4j 2.x will be upgraded to use Log4j 1.2.17 (or higher) and Log4j 2.17.1 (or higher) respectively, and will not require using the bootstrap actions provided above to mitigate the CVE issues.

- **Are EMR releases impacted by CVE-2021-45105?**

  The applications installed by Amazon EMR with EMR's default configurations are not impacted by CVE-2021-45105. Among applications installed by Amazon EMR, only Apache Hive uses Apache Log4j with context lookups, and it does not use non-default pattern layout in a manner that allows inappropriate input data to be processed.

- **Is Amazon EMR impacted by any of the following CVE disclosures?**

  The following table contains a list of CVEs that are related to Log4j and notes whether each CVE impacts Amazon EMR. The information in this table only applies when applications are installed by Amazon EMR using the default configurations.

CVE	Impacts EMR	Notes
CVE-2022-23302	No	Amazon EMR does not set up Log4j JMSSink
CVE-2022-23305	No	Amazon EMR does not set up Log4j JDBCAppender
CVE-2022-23307	No	Amazon EMR does not set up Log4j Chainsaw

CVE	Impacts EMR	Notes
CVE-2020-9493	No	Amazon EMR does not set up Log4j Chainsaw
CVE-2021-44832	No	Amazon EMR does not set up Log4j JDBCAppender with a JNDI connection string
CVE-2021-4104	No	Amazon EMR does not use Log4j JMSAppender
CVE-2020-9488	No	The applications that are installed by Amazon EMR do not use Log4j SMTPAppender
CVE-2019-17571	No	Amazon EMR blocks public access to clusters and does not launch SocketServer
CVE-2019-17531	No	We recommend that you upgrade to the latest Amazon EMR release version. Amazon EMR 5.33.0 and later use jackson-databind 2.6.7.4 or later, and EMR 6.1.0 and later use jackson-databind 2.10.0 or later. These versions of jackson-databind are not impacted by the CVE.

Amazon EMR archive of release notes

Release notes for all Amazon EMR releases are available below. For comprehensive release information for each release, see Amazon EMR 6.x release versions, Amazon EMR 5.x release versions and Amazon EMR 4.x release versions.

To get updates when a new Amazon EMR release is available, subscribe to the RSS feed for Amazon EMR release notes.

Release 6.14.0

The following release notes include information for Amazon EMR release 6.14.0. Changes are relative to 6.13.0. For information on the release timeline, see the 6.14.0 change log.

New features

- Amazon EMR 6.14.0 supports Apache Spark 3.4.1, Apache Spark RAPIDS 23.06.0-amzn-2, Flink 1.17.1, Iceberg 1.3.1, and Trino 422.
- Amazon EMR managed scaling is now available in the `ap-southeast-3` Asia Pacific (Jakarta) Region for clusters that you create with Amazon EMR 6.14.0 and higher.

Changes, enhancements, and resolved issues

- The 6.14.0 release optimizes log management with Amazon EMR running on Amazon EC2. As a result, you might see a slight reduction in storage costs for your cluster logs.
- The 6.14.0 release improves the scaling workflow to account for different core instances that have a substantial variation in size for their Amazon EBS volumes. This improvement applies to core nodes only; scale-down operations for task nodes aren't affected.
- The 6.14.0 release improves the way that Amazon EMR interacts with open-source applications such as Apache Hadoop YARN ResourceManager and HDFS NameNode. This improvement reduces the risk of operational delays with cluster scaling, and mitigates startup failures that occur due to connectivity issues with the open-source applications.
- The 6.14.0 release optimizes application installation at cluster launch. This improves the cluster startup times for certain combinations of Amazon EMR applications.
- The 6.14.0 release fixes an issue where cluster scale-down operations might stall when a cluster that's running in a VPC with a custom domain encounters a core or task node restart.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
			(Tel Aviv), Canada West (Calgary)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 223.0	4.14.336	March 8, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240131.0	4.14.336	February 14, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240124.0	4.14.336	February 7, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240109.0	4.14.334	January 24, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) ,Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 218.0	4.14.330	January 2, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 206.0	4.14.330	December 22, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231116.0	4.14.328	December 11, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 101.0	4.14.327	November 17, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230906.0	4.14.322	September 11, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

Release 6.13.0

The following release notes include information for Amazon EMR release 6.13.0. Changes are relative to 6.12.0. For information on the release timeline, see the 6.13.0 change log.

New features

- Amazon EMR 6.13.0 supports Apache Spark 3.4.1, Apache Spark RAPIDS 23.06.0-amzn-1, CUDA Toolkit 11.8.0, and JupyterHub 1.5.0.

Changes, enhancements, and resolved issues

- The 6.13.0 release improves the Amazon EMR log management daemon to ensure that all logs are uploaded at a regular cadence to Amazon S3 when a cluster termination command is issued. This facilitates faster cluster terminations.
- The 6.13.0 release enhances Amazon EMR log management capabilities to ensure consistent and timely upload of all log files to Amazon S3. This especially benefits long-running EMR clusters.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240816.0	4.14.350	August 21, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
			(Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 809.0	4.14.349	August 20, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240719.0	4.14.348	July 25, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
			(Tel Aviv), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240223.0	4.14.336	March 8, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240131.0	4.14.336	February 14, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 124.0	4.14.336	February 7, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240109.0	4.14.334	January 24, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023...218.0	4.14.330	January 2, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231206.0	4.14.330	December 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231116.0	4.14.328	December 11, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231101.0	4.14.327	November 16, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 020.1	4.14.326	November 7, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231012.1	4.14.326	October 26, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 926.0	4.14.322	October 19, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230906.0	4.14.322	October 4, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 808.0	4.14.320	August 24, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

Release 6.12.0

The following release notes include information for Amazon EMR release 6.12.0. Changes are relative to 6.11.0. For information on the release timeline, see the 6.12.0 change log.

New features

- Amazon EMR 6.12.0 supports Apache Spark 3.4.0, Apache Spark RAPIDS 23.06.0-amzn-0, CUDA 11.8.0, Apache Hudi 0.13.1-amzn-0, Apache Iceberg 1.3.0-amzn-0, Trino 414, and PrestoDB 0.281.
- Amazon EMR releases 6.12.0 and higher support LDAP integration with Apache Livy, Apache Hive through HiveServer2 (HS2), Trino, Presto, and Hue. You can also install Apache Spark and Apache Hadoop on an EMR cluster that uses 6.12.0 or higher and configure them to use LDAP. For more information, see Use Active Directory or LDAP servers for authentication with Amazon EMR.

Changes, enhancements, and resolved issues

- Amazon EMR releases 6.12.0 and higher provide Java 11 runtime support for Flink. For more information, see Configure Flink to run with Java 11.
- The 6.12.0 release adds a new retry mechanism to the cluster scaling workflow for EMR clusters that run Presto or Trino. This improvement reduces the risk that cluster resizing will indefinitely stall due to a single failed resize operation. It also improves cluster utilization, because your cluster scales up and down faster.
- The 6.12.0 release fixes an issue where cluster scale-down operations might stall when a core node that is undergoing graceful decommissioning turns unhealthy for any reason before it fully decommissions.
- The 6.12.0 release improves cluster scale-down logic so that your cluster doesn't attempt a scale-down of core nodes below the HDFS replication factor setting for the cluster. This aligns with your data redundancy requirements, and reduces the chance that a scaling operation might stall.
- The 6.12.0 release enhances the performance and efficiency of the health monitoring service for Amazon EMR by increasing the speed at which it logs state changes for instances. This improvement reduces the chance of degraded performance for cluster nodes that are running multiple custom client tools or third-party applications.
- The 6.12.0 release improves the performance of the on-cluster log management daemon for Amazon EMR. As a result, there is less chance for degraded performance with EMR clusters that run steps with high concurrency.
- With Amazon EMR release 6.12.0, the log management daemon has been upgraded to identify all logs that are in active use with open file handles on the local instance storage, and the associated processes. This upgrade ensures that Amazon EMR properly deletes the files and reclaims storage space after the logs are archived to Amazon S3.

- The 6.12.0 release includes a log-management daemon enhancement that deletes empty, unused steps directories in the local cluster file system. An excessively large number of empty directories can degrade the performance of Amazon EMR daemons and result in disk over-utilization.
- The 6.12.0 release enables log rotation for YARN Timeline Server logs. This minimizes disk over-utilization scenarios, especially for long-running clusters.
- The default root volume size has increased to 15 GB in Amazon EMR 6.10.0 and higher. Earlier releases have default root volume size of 10 GB.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240816.0	4.14.350	August 21, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
			Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 809.0	4.14.349	August 20, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 719.0	4.14.348	July 25, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
			(Tel Aviv), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 223.0	4.14.336	March 8, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240131.0	4.14.336	February 14, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240124.0	4.14.336	February 7, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240109.0	4.14.334	January 24, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 218.0	4.14.330	January 2, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231206.0	4.14.330	December 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231116.0	4.14.328	December 11, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 101.0	4.14.327	November 16, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231020.1	4.14.326	November 7, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20231012.1	4.14.326	October 26, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230926.0	4.14.322	October 19, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230906.0	4.14.322	October 4, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230822.0	4.14.322	Augest 30, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230808.0	4.14.320	August 24, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230727.0	4.14.320	August 14, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230719.0	4.14.320	August 2, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230628.0	4.14.318	July 12, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central)

Release 6.11.1

The following release notes include information for Amazon EMR release 6.11.1. Changes are relative to 6.11.0. For information on the release timeline, see the 6.11.1 change log.

Changes, enhancements, and resolved issues

- Due to lock contention, a node can enter into a deadlock if it's added or removed at the same time that it attempts to decommission. As a result, the Hadoop Resource Manager (YARN) becomes unresponsive, and affects all the incoming and currently-running containers.
- This release includes a change that allows high-availability clusters to recover from failed state after restart.
- This release includes security fixes for Hue and HBase.
- This release fixes an issue where clusters that are running workloads on Spark with Amazon EMR might silently receive incorrect results with `contains`, `startsWith`, `endsWith`, and `like`. This issue occurs when you use the expressions on partitioned fields that have metadata in the Amazon EMR Hive3 Metastore Server (HMS).
- This release fixes an issue with throttling on the Glue side when there are no user-defined functions (UDF).
- This release fixes an issue that deletes container logs by the node log aggregation service before log pusher can push them to S3 in case of YARN decommissioning.
- This release fixes an issue with FairShare Scheduler metrics when Node Label is enabled for Hadoop.
- This release fixes an issue that impacted Spark performance when you set a default `true` value for the `spark.yarn.heterogeneousExecutors.enabled` config in `spark-defaults.conf`.
- This release fixes an issue with Reduce Task failing to read shuffle data. The issue caused Hive query failures with a corrupted memory error.
- This release adds a new retry mechanism to the cluster scaling workflow for EMR clusters that run Presto or Trino. This improvement reduces the risk that cluster resizing will indefinitely stall due to a single failed resize operation. It also improves cluster utilization, because your cluster scales up and down faster.
- This release improves cluster scale-down logic so that your cluster doesn't attempt a scale-down of core nodes below the HDFS replication factor setting for the cluster. This aligns with your data redundancy requirements, and reduces the chance that a scaling operation might stall.
- The log management daemon has been upgraded to identify all logs that are in active use with open file handles on the local instance storage, and the associated processes. This upgrade ensures that Amazon EMR properly deletes the files and reclaims storage space after the logs are archived to Amazon S3.

- This release includes a log-management daemon enhancement that deletes empty, unused steps directories in the local cluster file system. An excessively large number of empty directories can degrade the performance of Amazon EMR daemons and result in disk over-utilization.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 816.0	4.14.350	August 21, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
			(US-East), China (Beijing), China (Ningxia)
2.0.20240809.0	4.14.349	August 20, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240719.0	4.14.348	July 25, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 223.0	4.14.336	March 8, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 131.0	4.14.336	February 14, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240124.0	4.14.336	February 7, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240109.0	4.14.334	January 24, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Canada West (Calgary)

OsRelea Label (Amazoı Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 218.0	4.14.330	January 2, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20231206.0	4.14.330	December 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023116.0	4.14.328	December 11, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 101.0	4.14.327	November 16, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central) , Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East) , China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 020.1	4.14.326	November 7, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023012.1	4.14.326	October 26, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230926.0	4.14.322	October 19, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 906.0	4.14.322	October 4, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 822.0	4.14.322	Augest 30, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230808.0	4.14.320	August 24, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsRelease Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 727.0	4.14.320	August 14, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

Release 6.11.0

The following release notes include information for Amazon EMR release 6.11.0. Changes are relative to 6.10.0. For information on the release timeline, see the change log.

New features

- Amazon EMR 6.11.0 supports Apache Spark 3.3.2-amzn-0, Apache Spark RAPIDS 23.02.0-amzn-0, CUDA 11.8.0, Apache Hudi 0.13.0-amzn-0, Apache Iceberg 1.2.0-amzn-0, Trino 410-amzn-0, and PrestoDB 0.279-amzn-0.

Changes, enhancements, and resolved issues

- With Amazon EMR 6.11.0, the DynamoDB connector has been upgraded to version 5.0.0. Version 5.0.0 uses AWS SDK for Java 2.x. Previous releases used AWS SDK for Java 1.x. As a result of this upgrade, we strongly advise you to test your code before you use the DynamoDB connector with Amazon EMR 6.11.
- When the DynamoDB connector for Amazon EMR 6.11.0 calls the DynamoDB service, it uses the Region value that you provide for the `dynamodb.endpoint` property. We recommend that you also configure `dynamodb.region` when you use `dynamodb.endpoint`, and that both properties target the same AWS Region. If you use `dynamodb.endpoint` and you don't configure `dynamodb.region`, the DynamoDB connector for Amazon EMR 6.11.0 will return an invalid Region exception and attempt to reconcile your AWS Region information from the Amazon EC2 instance metadata service (IMDS). If the connector can't retrieve the Region from IMDS, it defaults to US East (N. Virginia) (`us-east-1`). The following error is an example of the invalid Region exception that you might get if you don't properly configure the `dynamodb.region` property: `error software.amazon.awssdk.services.dynamodb.model.DynamoDbException: Credential should be scoped to a valid region.` For more information on the classes that are affected by the AWS SDK for Java upgrade to 2.x, see the [Upgrade AWS SDK for Java from 1.x to 2.x (#175)](#) commit in the GitHub repo for the Amazon EMR - DynamoDB connector.
- This release fixes an issue where column data becomes `NULL` when you use Delta Lake to store Delta table data in Amazon S3 after column rename operation. For more information about this experimental feature in Delta Lake, see [Column rename operation](#) in the Delta Lake User Guide.
- The 6.11.0 release fixes an issue that might occur when you create an edge node by replicating one of the primary nodes from a cluster with multiple primary nodes. The replicated edge node could cause delays with scale-down operations, or result in high memory-utilization on the primary nodes. For more information on how to create an edge node to communicate with your EMR cluster, see [Edge Node Creator](#) in the `aws-samples` repo on GitHub.

- The 6.11.0 release improves the automation process that Amazon EMR uses to re-mount Amazon EBS volumes to an instance after a reboot.
- The 6.11.0 release fixes an issue that resulted in intermittent gaps in the Hadoop metrics that Amazon EMR publishes to Amazon CloudWatch.
- The 6.11.0 release fixes an issue with EMR clusters where an update to the YARN configuration file that contains the exclusion list of nodes for the cluster is interrupted due to disk over-utilization. The incomplete update hinders future cluster scale-down operations. This release ensures that your cluster remains healthy, and that scaling operations work as expected.
- The default root volume size has increased to 15 GB in Amazon EMR 6.10.0 and higher. Earlier releases have default root volume size of 10 GB.
- Hadoop 3.3.3 introduced a change in YARN (YARN-9608) that keeps nodes where containers ran in a decommissioning state until the application completes. This change ensures that local data such as shuffle data doesn't get lost, and you don' need to re-run the job. This approach might also lead to underutilization of resources on clusters with or without managed scaling enabled.

  With Amazon EMR releases 6.11.0 and higher as well as 6.8.1, 6.9.1, and 6.10.1, the value of `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` is set to `false` in `yarn-site.xml` to resolve this issue.

  While the fix addresses the issues that were introduced by YARN-9608, it might cause Hive jobs to fail due to shuffle data loss on clusters that have managed scaling enabled. We've mitigated that risk in this release by also setting `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` for Hive workloads. This config is only available with Amazon EMR releases 6.11.0 and higher.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

Note

This release no longer gets automatic AMI updates since it has been succeeded by 1 more more patch releases. The patch release is denoted by the number after the second decimal point (`6.8.`*1*). To see if you're using the latest patch release, check the available releases in the *Release Guide*, or check the **Amazon EMR release** dropdown when you create a cluster in the console, or use the `ListReleaseLabels` API or `list-release-`

labels CLI action. To get updates about new releases, subscribe to the RSS feed on the What's new? page.

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240816.0	4.14.350	August 21, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240809.0	4.14.349	August 20, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240719.0	4.14.348	July 25, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
			(Tel Aviv), Canada West (Calgary)
2.0.20230808.0	4.14.320	August 24, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 727.0	4.14.320	August 14, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 719.0	4.14.320	August 2, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230628.0	4.14.318	July 12, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Zurich), Europe (Milan), Europe (Spain), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 612.0	4.14.314	June 23, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Zurich), Europe (Milan), Europe (Spain), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderaba d), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230504.1	4.14.313	May 16, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central)

Release 6.10.0

The following release notes include information for Amazon EMR release 6.10.0. Changes are relative to 6.9.0. For information on the release timeline, see the change log.

New features

- Amazon EMR 6.10.0 supports Apache Spark 3.3.1, Apache Spark RAPIDS 22.12.0, CUDA 11.8.0, Apache Hudi 0.12.2-amzn-0, Apache Iceberg 1.1.0-amzn-0, Trino 403, and PrestoDB 0.278.1.
- Amazon EMR 6.10.0 includes a native Trino-Hudi connector that provides read access to data in Hudi tables. You can activate the connector with `trino-cli --catalog hudi`, and configure the connector for your requirements with `trino-connector-hudi`. The native integration with Amazon EMR means that you no longer need to use `trino-connector-hive` to query Hudi tables. For a list of supported configurations with the new connector, see the Hudi connector page of the Trino documentation.
- Amazon EMR releases 6.10.0 and higher support Apache Zeppelin integration with Apache Flink. See Working with Flink jobs from Zeppelin in Amazon EMR for more information.

Known Issues

- Hadoop 3.3.3 introduced a change in YARN (YARN-9608) that keeps nodes where containers ran in a decommissioning state until the application completes. This change ensures that local data such as shuffle data doesn't get lost, and you don' need to re-run the job. This approach might also lead to underutilization of resources on clusters with or without managed scaling enabled.

  To work around this issue in Amazon EMR 6.10.0, you can set the value of `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` to `false` in `yarn-site.xml`. In Amazon EMR releases 6.11.0 and higher as well as 6.8.1, 6.9.1, and 6.10.1, the config is set to `false` by default to resolve this issue.

Changes, enhancements, and resolved issues

- Amazon EMR 6.10.0 removes the dependency on `minimal-json.jar` for the Amazon Redshift integration for Apache Spark, and automatically adds the required Spark-Redshift related jars to the executor class path for Spark: `spark-redshift.jar`, `spark-avro.jar`, and `RedshiftJDBC.jar`.
- The 6.10.0 release improves the on-cluster log management daemon to monitor additional log folders in your EMR cluster. This improvement minimizes disk over-utilization scenarios.
- The 6.10.0 release automatically restarts the on-cluster log management daemon when it stops. This improvement reduces the risk for nodes to appear unhealthy due to disk over-utilization.
- Amazon EMR 6.10.0 supports regional endpoints for EMRFS user mapping.

- The default root volume size has increased to 15 GB in Amazon EMR 6.10.0 and higher. Earlier releases have default root volume size of 10 GB.
- The 6.10.0 release fixes an issue that caused Spark jobs to stall when all remaining Spark executors are on a decommissioning host with the YARN resource manager.
- With Amazon EMR 6.6.0 through 6.9.x, INSERT queries with dynamic partition and an ORDER BY or SORT BY clause will always have two reducers. This issue is caused by OSS change HIVE-20703, which puts dynamic sort partition optimization under cost-based decision. If your workload doesn't require sorting of dynamic partitions, we recommend that you set the `hive.optimize.sort.dynamic.partition.threshold` property to `-1` to disable the new feature and get the correctly calculated number of reducers. This issue is fixed in OSS Hive as part of HIVE-22269 and is fixed in Amazon EMR 6.10.0.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

> **Note**
>
> This release no longer gets automatic AMI updates since it has been succeeded by 1 more more patch releases. The patch release is denoted by the number after the second decimal point (`6.8.1`). To see if you're using the latest patch release, check the available releases in the *Release Guide*, or check the **Amazon EMR release** dropdown when you create a cluster in the console, or use the `ListReleaseLabels` API or `list-release-labels` CLI action. To get updates about new releases, subscribe to the RSS feed on the What's new? page.

OsRelease Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240816.0	4.14.350	August 21, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
			West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 809.0	4.14.349	August 20, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240719.0	4.14.348	July 25, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 808.0	4.14.320	August 24, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230727.0	4.14.320	August 14, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 719.0	4.14.320	August 2, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 628.0	4.14.318	July 12, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Zurich), Europe (Milan), Europe (Spain), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 612.0	4.14.314	June 23, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Zurich), Europe (Milan), Europe (Spain), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230504.1	4.14.313	May 16, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Zurich), Europe (Milan), Europe (Spain), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230418.0	4.14.311	May 3, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Zurich), Europe (Milan), Europe (Spain), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230404.1	4.14.311	April 18, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE)
2.0.20230404.0	4.14.311	April 10, 2023	US East (N. Virginia), Europe (Paris)

OsRelea Label (Amazo Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 320.0	4.14.309	March 30, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 207.0	4.14.304	February 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE)

Release 6.9.0

The following release notes include information for Amazon EMR release 6.9.0. Changes are relative to Amazon EMR release 6.8.0. For information on the release timeline, see the change log.

New Features

- Amazon EMR release 6.9.0 supports Apache Spark RAPIDS 22.08.0, Apache Hudi 0.12.1, Apache Iceberg 0.14.1, Trino 398, and Tez 0.10.2.
- Amazon EMR release 6.9.0 includes a new open-source application, Delta Lake 2.1.0.
- The Amazon Redshift integration for Apache Spark is included in Amazon EMR releases 6.9.0 and later. Previously an open-source tool, the native integration is a Spark connector that you can use to build Apache Spark applications that read from and write to data in Amazon Redshift and Amazon Redshift Serverless. For more information, see Using Amazon Redshift integration for Apache Spark with Amazon EMR .
- Amazon EMR release 6.9.0 adds support for archiving logs to Amazon S3 during cluster scale-down. Previously, you could only archive log files to Amazon S3 during cluster termination. The new capability ensures that log files generated on the cluster persist on Amazon S3 even after the node is terminated. For more information, see Configure cluster logging and debugging.
- To support long running queries, Trino now includes a fault-tolerant execution mechanism. Fault-tolerant execution mitigates query failures by retrying failed queries or their component tasks. For more information, see Fault-tolerant execution in Trino.
- You can use Apache Flink on Amazon EMR for unified `BATCH` and `STREAM` processing of Apache Hive Tables or metadata of any Flink tablesource such as Iceberg, Kinesis or Kafka. You can specify the AWS Glue Data Catalog as the metastore for Flink using the AWS Management Console, AWS CLI, or Amazon EMR API. For more information, see Configuring Flink in Amazon EMR.
- You can now specify AWS Identity and Access Management (IAM) runtime roles and AWS Lake Formation-based access control for Apache Spark, Apache Hive, and Presto queries on Amazon EMR on EC2 clusters with Amazon SageMaker Studio. For more information, see Configure runtime roles for Amazon EMR steps.

Known Issues

- For Amazon EMR release 6.9.0, Trino does not work on clusters enabled for Apache Ranger. If you need to use Trino with Ranger, contact AWS Support.
- If you use the the Amazon Redshift integration for Apache Spark and have a time, timetz, timestamp, or timestamptz with microsecond precision in Parquet format, the connector rounds the time values to the nearest millisecond value. As a workaround, use the text unload format `unload_s3_format` parameter.

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&‘()*+,-`. For more information, see [UTF-8 encoding table and Unicode characters](#).

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.
- Connections to Amazon EMR clusters from Amazon SageMaker Studio may intermittently fail with a **403 Forbidden** response code. This error happens when setup of the IAM role on the cluster takes longer than 60 seconds. As a workaround, you can install an Amazon EMR patch to enable retries and increase the timeout to a minimum of 300 seconds. Use the following steps to apply the bootstrap action when you launch your cluster.

  1. Download the bootstrap script and RPM files from the following Amazon S3 URIs.

     ```
     s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/gcsc/replace-rpms.sh
     s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/gcsc/emr-secret-agent-1.18.0-SNAPSHOT20221121212949.noarch.rpm
     ```
  2. Upload the files from the previous step to an Amazon S3 bucket that you own. The bucket must be in the same AWS Region where you plan to launch the cluster.
  3. Include the following bootstrap action when you launch your EMR cluster. Replace *`bootstrap_URI`* and *`RPM_URI`* with the corresponding URIs from Amazon S3.

     ```
     --bootstrap-actions "Path=bootstrap_URI,Args=[RPM_URI]"
     ```
- With Amazon EMR releases 5.36.0 and 6.6.0 through 6.9.0, `SecretAgent` and `RecordServer` service components may experience log data loss due to an incorrect file name pattern

configuration in Log4j2 properties. The incorrect configuration causes the components to generate only one log file per day. When the rotation strategy occurs, it overwrites the existing file instead of generating a new log file as expected. As a workaround, use a bootstrap action to generate log files each hour and append an auto-increment integer in the file name to handle the rotation.

For Amazon EMR 6.6.0 through 6.9.0 releases, use the following bootstrap action when you launch a cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-
actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

For Amazon EMR 5.36.0, use the following bootstrap action when you launch a cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-
actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

- Apache Flink provides Native S3 FileSystem and Hadoop FileSystem Connectors, which let applications create a FileSink and write the data into Amazon S3. This FileSink fails with one of the following two exceptions.

```
java.lang.UnsupportedOperationException: Recoverable writers on Hadoop are only
 supported for HDFS
```

```
Caused by: java.lang.NoSuchMethodError:
 org.apache.hadoop.io.retry.RetryPolicies.retryOtherThanRemoteAndSaslException(Lorg/
apache/hadoop/io/retry/RetryPolicy;Ljava/util/Map;)Lorg/apache/hadoop/io/retry/
RetryPolicy;
                                        at
 org.apache.hadoop.yarn.client.RMProxy.createRetryPolicy(RMProxy.java:302) ~[hadoop-
yarn-common-3.3.3-amzn-0.jar:?]
```

As a workaround, you can install an Amazon EMR patch, which fixes the above issue in Flink. To apply the bootstrap action when you launch your cluster, complete the following steps.

1. Download the flink-rpm to your Amazon S3 bucket. Your RPM path is `s3://DOC-EXAMPLE-BUCKET/rpms/flink/`.
2. Download the bootstrap script and RPM files from Amazon S3 using the following URI. Replace *regionName* with the AWS Region where you plan to launch the cluster.

```
s3://emr-data-access-control-regionName/customer-bootstrap-actions/gcsc/replace-
rpms.sh
```

3. Hadoop 3.3.3 introduced a change in YARN (YARN-9608) that keeps nodes where containers ran in a decommissioning state until the application completes. This change ensures that local data such as shuffle data doesn't get lost, and you don' need to re-run the job. In Amazon EMR 6.8.0 and 6.9.0, this approach might also lead to underutilization of resources on clusters with or without managed scaling enabled.

   With Amazon EMR 6.10.0, there's a workaround for this issue to set the value of `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` to `false` in `yarn-site.xml`. In Amazon EMR releases 6.11.0 and higher as well as 6.8.1, 6.9.1, and 6.10.1, the config is set to `false` by default to resolve this issue.

Changes, Enhancements, and Resolved Issues

- For Amazon EMR release 6.9.0 and later, all components installed by Amazon EMR that use Log4j libraries use Log4j version 2.17.1 or later.
- When you use the DynamoDB connector with Spark on Amazon EMR versions 6.6.0, 6.7.0, and 6.8.0, all reads from your table return an empty result, even though the input split references non-empty data. Amazon EMR release 6.9.0 fixes this issue.
- Amazon EMR 6.9.0 adds limited support for Lake Formation-based access control with Apache Hudi when reading data using Spark SQL. The support is for SELECT queries using Spark SQL and is limited to column-level access control. For more information, see Hudi and Lake Formation.
- When you use Amazon EMR 6.9.0 to create a Hadoop cluster with Node Labels enabled, the YARN metrics API returns aggregated information across all partitions, instead of the default partition. For more information, see YARN-11414.
- With Amazon EMR release 6.9.0, we've updated Trino to version 398, which uses Java 17. The previous supported version of Trino for Amazon EMR 6.8.0 was Trino 388 running on Java 11. For more information about this change, see Trino updates to Java 17 on the Trino blog.
- This releases fixes a timing sequence mismatch issue between Apache BigTop and the Amazon EMR on EC2 cluster startup sequence. This timing sequence mismatch occurs when a system attempts to perform two or more operations at the same time instead of doing them in the proper sequence. As a result, certain cluster configurations experienced instance startup timeouts and slower cluster startup times.

- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

> **Note**
>
> This release no longer gets automatic AMI updates since it has been succeeded by 1 more more patch releases. The patch release is denoted by the number after the second decimal point (`6.8.1`). To see if you're using the latest patch release, check the available releases in the *Release Guide*, or check the **Amazon EMR release** dropdown when you create a cluster in the console, or use the `ListReleaseLabels` API or `list-release-labels` CLI action. To get updates about new releases, subscribe to the RSS feed on the What's new? page.

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240816.0	4.14.350	August 21, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2024 809.0	4.14.349	August 20, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240719.0	4.14.348	July 25, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20240709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv), Canada West (Calgary)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023808.0	4.14.320	August 24, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), Israel (Tel Aviv)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 727.0	4.14.320	August 14, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfur t), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230719.0	4.14.320	August 2, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230628.0	4.14.318	July 12, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 612.0	4.14.314	June 23, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230504.1	4.14.313	May 16, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 418.0	4.14.311	May 3, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023 404.1	4.14.311	April 18, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)
2.0.2023 404.0	4.14.311	April 10, 2023	US East (N. Virginia), Europe (Paris)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2023320.0	4.14.309	March 30, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230307.0	4.14.305	March 15, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20230207.0	4.14.304	February 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.20221210.1	4.14.301	January 12, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazon Linux version)	Amazon Linux kernel version	Available date	Supported Regions
2.0.2022 103.3	4.14.296	December 5, 2022	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

Release 6.8.0

The following release notes include information for Amazon EMR release 6.8.0. Changes are relative to 6.7.0.

New Features

- Amazon EMR steps feature now supports Apache Livy endpoint and JDBC/ODBC clients. For more information, see Configure runtime roles for Amazon EMR steps.

- Amazon EMR release 6.8.0 comes with Apache HBase release 2.4.12. With this HBase release, you can both archive and delete your HBase tables. The Amazon S3 archive process renames all table files to the archive directory. This can be a costly and lengthy process. Now, you can skip the archive process and quickly drop and delete large tables. For more information, see Using the HBase shell.

Known Issues

- Hadoop 3.3.3 introduced a change in YARN (YARN-9608) that keeps nodes where containers ran in a decommissioning state until the application completes. This change ensures that local data such as shuffle data doesn't get lost, and you don' need to re-run the job. In Amazon EMR 6.8.0 and 6.9.0, this approach might also lead to underutilization of resources on clusters with or without managed scaling enabled.

  With Amazon EMR 6.10.0, there's a workaround for this issue to set the value of `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` to `false` in `yarn-site.xml`. In Amazon EMR releases 6.11.0 and higher as well as 6.8.1, 6.9.1, and 6.10.1, the config is set to `false` by default to resolve this issue.

Changes, Enhancements, and Resolved Issues

- When Amazon EMR release 6.5.0, 6.6.0, or 6.7.0 read Apache Phoenix tables through the Apache Spark shell, Amazon EMR produced a `NoSuchMethodError`. Amazon EMR release 6.8.0 fixes this issue.
- Amazon EMR release 6.8.0 comes with Apache Hudi 0.11.1; however, Amazon EMR 6.8.0 clusters are also compatible with the open-source `hudi-spark3.3-bundle_2.12` from Hudi 0.12.0.
- Amazon EMR release 6.8.0 comes with Apache Spark 3.3.0. This Spark release uses Apache Log4j 2 and the `log4j2.properties` file to configure Log4j in Spark processes. If you use Spark in the cluster or create EMR clusters with custom configuration parameters, and you want to upgrade to Amazon EMR release 6.8.0, you must migrate to the new `spark-log4j2` configuration classification and key format for Apache Log4j 2. For more information, see Migrating from Apache Log4j 1.x to Log4j 2.x.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

> **Note**
>
> This release no longer gets automatic AMI updates since it has been succeeded by 1 more more patch releases. The patch release is denoted by the number after the second decimal point (`6.8.1`). To see if you're using the latest patch release, check the available releases in the *Release Guide*, or check the **Amazon EMR release** dropdown when you create a cluster in the console, or use the `ListReleaseLabels` API or `list-release-labels` CLI action. To get updates about new releases, subscribe to the RSS feed on the What's new? page.

OsRelease Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 816.0	4.14.350	August 21, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 809.0	4.14.349	August 20, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazoı Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 719.0	4.14.348	July 25, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich), Asia Pacific (Melbourne), Israel (Tel Aviv)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230808.0	4.14.320	August 24, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

OsRelease Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230727.0	4.14.320	August 14, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central),

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230719.0	4.14.320	August 2, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Asia Pacific (Melbourne), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 628.0	4.14.318	July 12, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230612.0	4.14.314	June 23, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 504.1	4.14.313	May 16, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023[illegible] 418.0	4.14.311	May 3, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 404.1	4.14.311	April 18, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)
2.0.2023 404.0	4.14.311	April 10, 2023	US East (N. Virginia), Europe (Paris)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230320.0	4.14.309	March 30, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230307.0	4.14.305	March 15, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 207.0	4.14.304	February 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230119.1	4.14.301	February 3, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 210.1	4.14.301	December 22, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 103.3	4.14.296	December 5, 2022	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazoı Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 004.0	4.14.294	November 2, 2022	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelease Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20220912.1	4.14.291	September 6, 2022	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

Known Issues

- When you use the DynamoDB connector with Spark on Amazon EMR versions 6.6.0, 6.7.0, and 6.8.0, all reads from your table return an empty result, even though the input split references non-empty data. This is because Spark 3.2.0 sets `spark.hadoopRDD.ignoreEmptySplits` to `true` by default. As a workaround, explicitly set `spark.hadoopRDD.ignoreEmptySplits` to `false`. Amazon EMR release 6.9.0 fixes this issue.

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&‘()*+,-`. For more information, see [UTF-8 encoding table and Unicode characters](https://www.utf8-chartable.de/).

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

- With Amazon EMR releases 5.36.0 and 6.6.0 through 6.9.0, `SecretAgent` and `RecordServer` service components may experience log data loss due to an incorrect file name pattern configuration in Log4j2 properties. The incorrect configuration causes the components to generate only one log file per day. When the rotation strategy occurs, it overwrites the existing file instead of generating a new log file as expected. As a workaround, use a bootstrap action to generate log files each hour and append an auto-increment integer in the file name to handle the rotation.

  For Amazon EMR 6.6.0 through 6.9.0 releases, use the following bootstrap action when you launch a cluster.

  ```
  --bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
  ```

  For Amazon EMR 5.36.0, use the following bootstrap action when you launch a cluster.

  ```
  --bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
  ```

For more information on the release timeline, see the [change log](#).

Release 6.7.0

The following release notes include information for Amazon EMR release 6.7.0. Changes are relative to 6.6.0.

Initial release date: July 15, 2022

New Features

- Amazon EMR now supports Apache Spark 3.2.1, Apache Hive 3.1.3, HUDI 0.11, PrestoDB 0.272, and Trino 0.378.
- Supports IAM Role and Lake Formation-based access controls with EMR steps (Spark, Hive) for Amazon EMR on EC2 clusters.
- Supports Apache Spark data definition statements on Apache Ranger enabled clusters. This now includes support for Trino applications reading and writing Apache Hive metadata on Apache Ranger enabled clusters. For more information, see Enable federated governance using Trino and Apache Ranger on Amazon EMR.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 816.0	4.14.350	August 21, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
			(Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 809.0	4.14.349	August 20, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 719.0	4.14.348	July 25, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 709.1	4.14.348	July 23, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia), Asia Pacific (Hyderabad), Middle East (UAE), Europe (Spain), Europe (Zurich)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 223.0	4.14.336	March 8, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 131.0	4.14.336	February 14, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240124.0	4.14.336	February 7, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 109.0	4.14.334	January 24, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 218.0	4.14.330	January 2, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazoı Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 206.0	4.14.330	December 22, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 116.0	4.14.328	December 11, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 101.0	4.14.327	November 16, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023...020.1	4.14.326	November 7, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20231012.1	4.14.326	October 26, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 926.0	4.14.322	October 19, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 906.0	4.14.322	October 4, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230822.0	4.14.322	August 30, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 808.0	4.14.320	August 24, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230727.0	4.14.320	August 14, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230719.0	4.14.320	August 2, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 628.0	4.14.318	July 12, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230612.0	4.14.314	June 23, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230504.1	4.14.313	May 16, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 418.0	4.14.311	May 3, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023404.1	4.14.311	April 18, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)
2.0.2023404.0	4.14.311	April 10, 2023	US East (N. Virginia), Europe (Paris)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230320.0	4.14.309	March 30, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230307.0	4.14.305	March 15, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 207.0	4.14.304	February 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 119.1	4.14.301	February 3, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 210.1	4.14.301	December 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 103.3	4.14.296	December 5, 2022	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20221004.0	4.14.294	November 2, 2022	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 912.1	4.14.291	October 7, 2022	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)
2.0.2022 719.0	4.14.287	August 10, 2022	`us-west-1` , `eu-west-3` , `eu-north-1` , `ap-south-1` , `me-south-1`

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 606.1	4.14.281	July 15, 2022	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

Known Issues

- When Amazon EMR release 6.5.0, 6.6.0, or 6.7.0 read Apache Phoenix tables through the Apache Spark shell, a `NoSuchMethodError` occurs because Amazon EMR uses an incorrect `Hbase.compat.version`. Amazon EMR release 6.8.0 fixes this issue.
- When you use the DynamoDB connector with Spark on Amazon EMR versions 6.6.0, 6.7.0, and 6.8.0, all reads from your table return an empty result, even though the input split references non-empty data. This is because Spark 3.2.0 sets `spark.hadoopRDD.ignoreEmptySplits` to

true by default. As a workaround, explicitly set spark.hadoopRDD.ignoreEmptySplits to false. Amazon EMR release 6.9.0 fixes this issue.

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, s3://bucket/table/p=a is a prefix of s3://bucket/table/p=a b.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in s3://bucket/table/p=a b falls into this category. Note that there are 14 other non-control characters: !"#$%&‘()*+,-. For more information, see [UTF-8 encoding table and Unicode characters](https://www.utf8-chartable.de/).

  As a workaround to this issue, set the spark.sql.sources.fastS3PartitionDiscovery.enabled configuration to false in the spark-defaults classification.

- With Amazon EMR releases 5.36.0 and 6.6.0 through 6.9.0, SecretAgent and RecordServer service components may experience log data loss due to an incorrect file name pattern configuration in Log4j2 properties. The incorrect configuration causes the components to generate only one log file per day. When the rotation strategy occurs, it overwrites the existing file instead of generating a new log file as expected. As a workaround, use a bootstrap action to generate log files each hour and append an auto-increment integer in the file name to handle the rotation.

  For Amazon EMR 6.6.0 through 6.9.0 releases, use the following bootstrap action when you launch a cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-
actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

  For Amazon EMR 5.36.0, use the following bootstrap action when you launch a cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-
actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

- The GetClusterSessionCredentials API isn't supported with clusters that run on Amazon EMR 6.7 or lower.

Release 6.6.0

The following release notes include information for Amazon EMR release 6.6.0. Changes are relative to 6.5.0.

Initial release date: May 9, 2022

Updated documentation date: June 15, 2022

New Features

- Amazon EMR 6.6 now supports Apache Spark 3.2, Apache Spark RAPIDS 22.02, CUDA 11, Apache Hudi 0.10.1, Apache Iceberg 0.13, Trino 0.367 and PrestoDB 0.267.
- When you launch a cluster with *the latest patch release* of Amazon EMR 5.36 or higher, 6.6 or higher, or 7.0 or higher, Amazon EMR uses the latest Amazon Linux 2023 or Amazon Linux 2 release for the default Amazon EMR AMI. For more information, see Using the default Amazon Linux AMI for Amazon EMR.

OsRelease Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 816.0	4.14.350	August 21, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
			(Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 809.0	4.14.349	August 20, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 719.0	4.14.348	July 25, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 709.1	4.14.348	July 23, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Eu rope (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024 223.0	4.14.336	March 8, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2024[...]131.0	4.14.336	February 14, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240124.0	4.14.336	February 7, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20240109.0	4.14.334	January 24, 2024	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 218.0	4.14.330	January 2, 2024	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20231206.0	4.14.330	December 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023116.0	4.14.328	December 11, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 101.0	4.14.327	November 16, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 020.1	4.14.326	November 7, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfur t), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20231012.1	4.14.326	October 26, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 926.0	4.14.322	October 19, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central), AWS GovCloud (US-West), AWS GovCloud (US-East), China (Beijing), China (Ningxia)

OsRelease Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 906.0	4.14.322	October 4, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023...822.0	4.14.322	August 30, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230808.0	4.14.320	August 24, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230727.0	4.14.320	August 14, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Canada (Central)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 719.0	4.14.320	August 2, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Europe (Stockholm), Europe (Milan), Europe (Spain), Europe (Frankfurt), Europe (Zurich), Europe (Ireland), Europe (London), Europe (Paris), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Hyderabad), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Jakarta), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain), Middle East (UAE), Canada (Central)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 628.0	4.14.318	July 12, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 612.0	4.14.314	June 23, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230504.1	4.14.313	May 16, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 418.0	4.14.311	May 3, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 404.1	4.14.311	April 18, 2023	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)
2.0.2023 404.0	4.14.311	April 10, 2023	US East (N. Virginia), Europe (Paris)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230320.0	4.14.309	March 30, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2023 307.0	4.14.305	March 15, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230207.0	4.14.304	February 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.20230119.1	4.14.301	February 3, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 210.1	4.14.301	December 22, 2023	US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 103.3	4.14.296	December 5, 2022	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 004.0	4.14.294	November 2, 2022	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 912.1	4.14.291	October 7, 2022	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)
2.0.2022 805.0	4.14.287	August 30, 2022	`us-west-1`

OsRelea Label (Amazoı Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 719.0	4.14.287	August 10, 2022	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsReleaseLabel (Amazon Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 426.0	4.14.281	June 10, 2022	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

OsRelea Label (Amazo Linux Version)	Amazon Linux Kernel Version	Available Date	Supported Regions
2.0.2022 406.1	4.14.275	May 2, 2022	US East (N. Virginia) , US East (Ohio), US West (N. California), US West (Oregon), Canada (Central), Europe (Stockholm), Europe (Ireland), Europe (London), Europe (Paris), Europe (Frankfurt), Europe (Milan), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Jakarta), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Osaka), Asia Pacific (Singapore), Asia Pacific (Sydney), Africa (Cape Town), South America (São Paulo), Middle East (Bahrain)

- With Amazon EMR 6.6 and later, applications that use Log4j 1.x and Log4j 2.x are upgraded to use Log4j 1.2.17 (or higher) and Log4j 2.17.1 (or higher) respectively, and do not require using the bootstrap actions provided to mitigate the CVE issues.
- **[Managed scaling] Spark shuffle data managed scaling optimization** - For Amazon EMR versions 5.34.0 and later, and EMR versions 6.4.0 and later, managed scaling is now Spark shuffle data aware (data that Spark redistributes across partitions to perform specific operations). For more information on shuffle operations, see Using EMR managed scaling in Amazon EMR in the *Amazon EMR Management Guide* and Spark Programming Guide.

- Starting with Amazon EMR 5.32.0 and 6.5.0, dynamic executor sizing for Apache Spark is enabled by default. To turn this feature on or off, you can use the `spark.yarn.heterogeneousExecutors.enabled` configuration parameter.

Changes, Enhancements, and Resolved Issues

- Amazon EMR reduces cluster startup time by up to 80 seconds on average for clusters that use the EMR default AMI option and only install common applications, such as Apache Hadoop, Apache Spark and Apache Hive.

Known Issues

- When Amazon EMR release 6.5.0, 6.6.0, or 6.7.0 read Apache Phoenix tables through the Apache Spark shell, a `NoSuchMethodError` occurs because Amazon EMR uses an incorrect `Hbase.compat.version`. Amazon EMR release 6.8.0 fixes this issue.
- When you use the DynamoDB connector with Spark on Amazon EMR versions 6.6.0, 6.7.0, and 6.8.0, all reads from your table return an empty result, even though the input split references non-empty data. This is because Spark 3.2.0 sets `spark.hadoopRDD.ignoreEmptySplits` to `true` by default. As a workaround, explicitly set `spark.hadoopRDD.ignoreEmptySplits` to `false`. Amazon EMR release 6.9.0 fixes this issue.
- On Trino long-running clusters, Amazon EMR 6.6.0 enables Garbage Collection logging parameters in the Trino jvm.config to get better insights from the Garbage Collection logs. This change appends many Garbage Collection logs to the launcher.log (/var/log/trino/launcher.log) file. If you are running Trino clusters in Amazon EMR 6.6.0, you may encounter nodes running out of disk space after the cluster has been running for a couple of days due to the appended logs.

  The workaround for this issue is to run the script below as a Bootstrap Action to disable the Garbage Collection logging parameters in jvm.config while creating or cloning the cluster for Amazon EMR 6.6.0.

```
#!/bin/bash
  set -ex
  PRESTO_PUPPET_DIR='/var/aws/emr/bigtop-deploy/puppet/modules/trino'
  sudo bash -c "sed -i '/-Xlog/d' ${PRESTO_PUPPET_DIR}/templates/jvm.config"
```

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:

- Two or more partitions are scanned from the same table.
- At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
- The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&'()*+,-`. For more information, see [UTF-8 encoding table and Unicode characters](#).

As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

- With Amazon EMR releases 5.36.0 and 6.6.0 through 6.9.0, `SecretAgent` and `RecordServer` service components may experience log data loss due to an incorrect file name pattern configuration in Log4j2 properties. The incorrect configuration causes the components to generate only one log file per day. When the rotation strategy occurs, it overwrites the existing file instead of generating a new log file as expected. As a workaround, use a bootstrap action to generate log files each hour and append an auto-increment integer in the file name to handle the rotation.

  For Amazon EMR 6.6.0 through 6.9.0 releases, use the following bootstrap action when you launch a cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

  For Amazon EMR 5.36.0, use the following bootstrap action when you launch a cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

Release 5.35.0

This is the Amazon EMR release 5.35.0 release note.

The following release notes include information for Amazon EMR release 5.35.0. Changes are relative to 5.34.0.

Initial release date: March 30, 2022

New Features

- Amazon EMR release 5.35 applications that use Log4j 1.x and Log4j 2.x are upgraded to use Log4j 1.2.17 (or higher) and Log4j 2.17.1 (or higher) respectively, and do not require using bootstrap actions to mitigate the CVE issues in previous releases. See Approach to mitigate CVE-2021-44228.

Changes, Enhancements, and Resolved Issues

Flink changes

Change type	Description
Upgrades	• Update flink version to 1.14.2. • log4j upgraded to 2.17.1.

Hadoop changes

Change type	Description
Hadoop open source backports since EMR 5.34.0	• YARN-10438: Handle null containerId in ClientRMService#getContainerReport() • YARN-7266: Timeline Server event handler threads locked • YARN-10438: ATS 1.5 fails to start if RollingLevelDb files are corrupt or missing • HADOOP-13500: Synchronizing iteration of Configuration properties object • YARN-10651: CapacityScheduler crashed with NPE in AbstractYarnScheduler.updat eNodeResource()

Change type	Description
	• HDFS-12221: Replace xerces in XmlEditsV isitor • HDFS-16410: Insecure Xml parsing in OfflineEditsXmlLoader
Hadoop changes and fixes	• Tomcat used in KMS and HttpFS is upgraded to 8.5.75 • In FileSystemOptimizedCommitterV2, the success marker was written in the commitJob output path defined while creating the committer. Since commitJob and task level output paths can differ, the path has been corrected to use the one defined in manifest files. For Hive jobs, this results in the success marker being written correctly in when performing operations such as dynamic partition or UNION ALL.

Hive changes

Change type	Description
Hive upgraded to open source release 2.3.9, including these JIRA fixes	• HIVE-17155: findConfFile() in HiveConf.java has some issues with the conf path • HIVE-24797: Disable validate default values when parsing Avro schemas • HIVE-21563: Improve Table#getEmptyTabl e performance by disable registerAllFunctio nsOnce • HIVE-18147: Tests can fail with java.net. BindException: Address already in use • HIVE-24608: Switch back to get_table in HMS client for Hive 2.3.x

Change type	Description
	• HIVE-21200: Vectorization - date column throwing java.lang.UnsupportedOperationException for parquet • HIVE-19228: Remove commons-httpclient 3.x usage
Hive open source backports since EMR 5.34.0	• HIVE-19990: Query with interval literal in join condition fails • HIVE-25824: Upgrade branch-2.3 to log4j 2.17.0 • TEZ-4062: Speculative attempt scheduling should be aborted when Task has completed • TEZ-4108: NullPointerException during speculative execution race condition • TEZ-3918: Setting tez.task.log.level does not work
Hive upgrades and fixes	• Upgrade Log4j version to 2.17.1 • Upgrade ORC version to 1.4.3 • FixED deadlock due to penalty thread in ShuffleScheduler
New features	• Added feature to print Hive Query in AM logs. This is disabled by default. Flag/Conf : `tez.am.emr.print.hive.query.in.log` . Status (default): FALSE.

Oozie changes

Change type	Description
Oozie open source backports since EMR 5.34.0	• OOZIE-3652: Oozie launcher should retry directory listing when NoSuchFileException occurs

Pig changes

Change type	Description
Upgrades	• log4j upgraded to 1.2.17.

Known issues

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&'()*+,-`. For more information, see [UTF-8 encoding table and Unicode characters](#).

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

Release 5.34.0

The following release notes include information for Amazon EMR release 5.34.0. Changes are relative to 5.33.1.

Initial release date: January 20, 2022

Updated release date: March 21, 2022

New Features

- **[Managed scaling] Spark shuffle data managed scaling optimization** - For Amazon EMR versions 5.34.0 and later, and EMR versions 6.4.0 and later, managed scaling is now Spark shuffle data aware (data that Spark redistributes across partitions to perform specific operations). For more information on shuffle operations, see Using EMR managed scaling in Amazon EMR in the *Amazon EMR Management Guide* and Spark Programming Guide.

- [Hudi] Improvements to simplify Hudi configuration. Disabled optimistic concurrency control by default.

Changes, Enhancements, and Resolved Issues

- This is a release to fix issues with Amazon EMR Scaling when it fails to scale up/scale down a cluster successfully or causes application failures.
- Previously, manual restart of the resource manager on a multi-master cluster caused Amazon EMR on-cluster daemons, like Zookeeper, to reload all previously decommissioned or lost nodes in the Zookeeper znode file. This caused default limits to be exceeded in certain situations. Amazon EMR now removes the decommissioned or lost node records older than one hour from the Zookeeper file and the internal limits have been increased.
- Fixed an issue where scaling requests failed for a large, highly utilized cluster when Amazon EMR on-cluster daemons were running health checking activities, such as gathering YARN node state and HDFS node state. This was happening because on-cluster daemons were not able to communicate the health status data of a node to internal Amazon EMR components.
- Improved EMR on-cluster daemons to correctly track the node states when IP addresses are reused to improve reliability during scaling operations.
- SPARK-29683. Fixed an issue where job failures occurred during cluster scale-down as Spark was assuming all available nodes were deny-listed.
- YARN-9011. Fixed an issue where job failures occurred due to a race condition in YARN decommissioning when cluster tried to scale up or down.
- Fixed issue with step or job failures during cluster scaling by ensuring that the node states are always consistent between the Amazon EMR on-cluster daemons and YARN/HDFS.
- Fixed an issue where cluster operations such as scale down and step submission failed for Amazon EMR clusters enabled with Kerberos authentication. This was because the Amazon EMR on-cluster daemon did not renew the Kerberos ticket, which is required to securely communicate with HDFS/YARN running on the primary node.
- Zeppelin upgraded to version 0.10.0.
- Livy Fix - upgraded to 0.7.1

- Spark performance improvement - heterogeneous executors are disabled when certain Spark configuration values are overridden in EMR 5.34.0.
- WebHDFS and HttpFS server are disabled by default. You can re-enable WebHDFS using the Hadoop configuration, `dfs.webhdfs.enabled`. HttpFS server can be started by using `sudo systemctl start hadoop-httpfs`.

Known Issues

- The Amazon EMR Notebooks feature used with Livy user impersonation does not work because HttpFS is disabled by default. In this case, the EMR notebook cannot connect to the cluster that has Livy impersonation enabled. The workaround is to start HttpFS server before connecting the EMR notebook to the cluster using `sudo systemctl start hadoop-httpfs`.
- Hue queries do not work in Amazon EMR 6.4.0 because Apache Hadoop HttpFS server is disabled by default. To use Hue on Amazon EMR 6.4.0, either manually start HttpFS server on the Amazon EMR primary node using `sudo systemctl start hadoop-httpfs`, or use an Amazon EMR step.
- The Amazon EMR Notebooks feature used with Livy user impersonation does not work because HttpFS is disabled by default. In this case, the EMR notebook cannot connect to the cluster that has Livy impersonation enabled. The workaround is to start HttpFS server before connecting the EMR notebook to the cluster using `sudo systemctl start hadoop-httpfs`.
- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that’s less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&‘()*+,-`. For more information, see UTF-8 encoding table and Unicode characters.

As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

Release 6.5.0

The following release notes include information for Amazon EMR release 6.5.0. Changes are relative to 6.4.0.

Initial release date: January 20, 2022

Updated release date: March 21, 2022

New Features

- **[Managed scaling] Spark shuffle data managed scaling optimization** - For Amazon EMR versions 5.34.0 and later, and EMR versions 6.4.0 and later, managed scaling is now Spark shuffle data aware (data that Spark redistributes across partitions to perform specific operations). For more information on shuffle operations, see Using EMR managed scaling in Amazon EMR in the *Amazon EMR Management Guide* and Spark Programming Guide.
- Starting with Amazon EMR 5.32.0 and 6.5.0, dynamic executor sizing for Apache Spark is enabled by default. To turn this feature on or off, you can use the `spark.yarn.heterogeneousExecutors.enabled` configuration parameter.
- Support for Apache Iceberg open table format for huge analytic datasets.
- Support for ranger-trino-plugin 2.0.1-amzn-1
- Support for toree 0.5.0

Changes, Enhancements, and Resolved Issues

- Amazon EMR 6.5 release version now supports Apache Iceberg 0.12.0, and provides runtime improvements with Amazon EMR Runtime for Apache Spark, Amazon EMR Runtime for Presto, and Amazon EMR Runtime for Apache Hive.
- Apache Iceberg is an open table format for large data sets in Amazon S3 and provides fast query performance over large tables, atomic commits, concurrent writes, and SQL-compatible table evolution. With EMR 6.5, you can use Apache Spark 3.1.2 with the Iceberg table format.

- Apache Hudi 0.9 adds Spark SQL DDL and DML support. This allows you to create, upsert Hudi tables using just SQL statements. Apache Hudi 0.9 also includes query side and writer side performance improvements.
- Amazon EMR Runtime for Apache Hive improves Apache Hive performance on Amazon S3 by removing rename operations during staging operations, and improves performance for metastore check (MSCK) commands used for repairing tables.

Known Issues

- When Amazon EMR release 6.5.0, 6.6.0, or 6.7.0 read Apache Phoenix tables through the Apache Spark shell, a `NoSuchMethodError` occurs because Amazon EMR uses an incorrect `Hbase.compat.version`. Amazon EMR release 6.8.0 fixes this issue.
- Hbase bundle clusters in high availability (HA) fail to provision with the default volume size and instance type. The workaround for this issue is to increase the root volume size.
- To use Spark actions with Apache Oozie, you must add the following configuration to your Oozie `workflow.xml` file. Otherwise, several critical libraries such as Hadoop and EMRFS will be missing from the classpath of the Spark executors that Oozie launches.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&‘()*+,-`. For more information, see [UTF-8 encoding table and Unicode characters](#).

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

Release 6.4.0

The following release notes include information for Amazon EMR release 6.4.0. Changes are relative to 6.3.0.

Initial release date: Sept 20, 2021

Updated release date: March 21, 2022

Supported applications

- AWS SDK for Java version 1.12.31
- CloudWatch Sink version 2.2.0
- DynamoDB Connector version 4.16.0
- EMRFS version 2.47.0
- Amazon EMR Goodies version 3.2.0
- Amazon EMR Kinesis Connector version 3.5.0
- Amazon EMR Record Server version 2.1.0
- Amazon EMR Scripts version 2.5.0
- Flink version 1.13.1
- Ganglia version 3.7.2
- AWS Glue Hive Metastore Client version 3.3.0
- Hadoop version 3.2.1-amzn-4
- HBase version 2.4.4-amzn-0
- HBase-operator-tools 1.1.0
- HCatalog version 3.1.2-amzn-5
- Hive version 3.1.2-amzn-5
- Hudi version 0.8.0-amzn-0
- Hue version 4.9.0
- Java JDK version Corretto-8.302.08.1 (build 1.8.0_302-b08)
- JupyterHub version 1.4.1
- Livy version 0.7.1-incubating
- MXNet version 1.8.0
- Oozie version 5.2.1

- Phoenix version 5.1.2
- Pig version 0.17.0
- Presto version 0.254.1-amzn-0
- Trino version 359
- Apache Ranger KMS (multi-master transparent encryption) version 2.0.0
- ranger-plugins 2.0.1-amzn-0
- ranger-s3-plugin 1.2.0
- SageMaker Spark SDK version 1.4.1
- Scala version 2.12.10 (OpenJDK 64-Bit Server VM, Java 1.8.0_282)
- Spark version 3.1.2-amzn-0
- spark-rapids 0.4.1
- Sqoop version 1.4.7
- TensorFlow version 2.4.1
- tez version 0.9.2
- Zeppelin version 0.9.0
- Zookeeper version 3.5.7
- Connectors and drivers: DynamoDB Connector 4.16.0

New features

- **[Managed scaling] Spark shuffle data managed scaling optimization** - For Amazon EMR versions 5.34.0 and later, and EMR versions 6.4.0 and later, managed scaling is now Spark shuffle data aware (data that Spark redistributes across partitions to perform specific operations). For more information on shuffle operations, see Using EMR managed scaling in Amazon EMR in the *Amazon EMR Management Guide* and Spark Programming Guide.
- On Apache Ranger-enabled Amazon EMR clusters, you can use Apache Spark SQL to insert data into or update the Apache Hive metastore tables using `INSERT INTO`, `INSERT OVERWRITE`, and `ALTER TABLE`. When using ALTER TABLE with Spark SQL, a partition location must be the child directory of a table location. Amazon EMR does not currently support inserting data into a partition where the partition location is different from the table location.
- PrestoSQL has been renamed to Trino.
- Hive: Execution of simple SELECT queries with LIMIT clause are accelerated by stopping the query execution as soon as the number of records mentioned in LIMIT clause is fetched. Simple

SELECT queries are queries that do not have GROUP BY / ORDER by clause or queries that do not have a reducer stage. For example, `SELECT * from <TABLE> WHERE <Condition> LIMIT <Number>`.

Hudi Concurrency Control

- Hudi now supports Optimistic Concurrency Control (OCC), which can be leveraged with write operations like UPSERT and INSERT to allow changes from multiple writers to the same Hudi table. This is file-level OCC, so any two commits (or writers) can write to the same table, if their changes do not conflict. For more information, see the Hudi concurrency control.
- Amazon EMR clusters have Zookeeper installed, which can be leveraged as the lock provider for OCC. To make it easier to use this feature, Amazon EMR clusters have the following properties pre-configured:

```
hoodie.write.lock.provider=org.apache.hudi.client.transaction.lock.ZookeeperBasedLockProvider
hoodie.write.lock.zookeeper.url=<EMR Zookeeper URL>
hoodie.write.lock.zookeeper.port=<EMR Zookeeper Port>
hoodie.write.lock.zookeeper.base_path=/hudi
```

  To enable OCC, you need to configure the following properties either with their Hudi job options or at the cluster-level using the Amazon EMR configurations API:

```
hoodie.write.concurrency.mode=optimistic_concurrency_control
hoodie.cleaner.policy.failed.writes=LAZY (Performs cleaning of failed writes lazily
 instead of inline with every write)
hoodie.write.lock.zookeeper.lock_key=<Key to uniquely identify the Hudi table> (Table
 Name is a good option)
```

Hudi Monitoring: Amazon CloudWatch integration to report Hudi Metrics

- Amazon EMR supports publishing Hudi Metrics to Amazon CloudWatch. It is enabled by setting the following required configurations:

```
hoodie.metrics.on=true
hoodie.metrics.reporter.type=CLOUDWATCH
```

- The following are optional Hudi configurations that you can change:

Setting	Description	Value
hoodie.metrics.cloudwatch.report.period.seconds	Frequency (in seconds) at which to report metrics to Amazon CloudWatch	Default value is 60s, which is fine for the default one minute resolution offered by Amazon CloudWatch
hoodie.metrics.cloudwatch.metric.prefix	Prefix to be added to each metric name	Default value is empty (no prefix)
hoodie.metrics.cloudwatch.namespace	Amazon CloudWatch namespace under which metrics are published	Default value is Hudi
hoodie.metrics.cloudwatch.maxDatumsPerRequest	Maximum number of datums to be included in one request to Amazon CloudWatch	Default value is 20, which is same as Amazon CloudWatch default

Amazon EMR Hudi configurations support and improvements

- Customers can now leverage EMR Configurations API and Reconfiguration feature to configure Hudi configurations at cluster level. A new file based configuration support has been introduced via /etc/hudi/conf/hudi-defaults.conf along the lines of other applications like Spark, Hive etc. EMR configures few defaults to improve user experience:

  — `hoodie.datasource.hive_sync.jdbcurl` is configured to the cluster Hive server URL and no longer needs to be specified. This is particularly useful when running a job in Spark cluster mode, where you previously had to specify the Amazon EMR master IP.

  — HBase specific configurations, which are useful for using HBase index with Hudi.

  — Zookeeper lock provider specific configuration, as discussed under concurrency control, which makes it easier to use Optimistic Concurrency Control (OCC).
- Additional changes have been introduced to reduce the number of configurations that you need to pass, and to infer automatically where possible:

  — The `partitionBy` keyword can be used to specify the partition column.

— When enabling Hive Sync, it is no longer mandatory to pass HIVE_TABLE_OPT_KEY, HIVE_PARTITION_FIELDS_OPT_KEY, HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY. Those values can be inferred from the Hudi table name and partition field.

— KEYGENERATOR_CLASS_OPT_KEY is not mandatory to pass, and can be inferred from simpler cases of SimpleKeyGenerator and ComplexKeyGenerator.

Hudi Caveats

- Hudi does not support vectorized execution in Hive for Merge on Read (MoR) and Bootstrap tables. For example, `count(*)` fails with Hudi realtime table when `hive.vectorized.execution.enabled` is set to true. As a workaround, you can disable vectorized reading by setting `hive.vectorized.execution.enabled` to `false`.
- Multi-writer support is not compatible with the Hudi bootstrap feature.
- Flink Streamer and Flink SQL are experimental features in this release. These features are not recommended for use in production deployments.

Changes, enhancements, and resolved issues

This is a release to fix issues with Amazon EMR Scaling when it fails to scale up/scale down a cluster successfully or causes application failures.

- Previously, manual restart of the resource manager on a multi-master cluster caused Amazon EMR on-cluster daemons, like Zookeeper, to reload all previously decommissioned or lost nodes in the Zookeeper znode file. This caused default limits to be exceeded in certain situations. Amazon EMR now removes the decommissioned or lost node records older than one hour from the Zookeeper file and the internal limits have been increased.
- Fixed an issue where scaling requests failed for a large, highly utilized cluster when Amazon EMR on-cluster daemons were running health checking activities, such as gathering YARN node state and HDFS node state. This was happening because on-cluster daemons were not able to communicate the health status data of a node to internal Amazon EMR components.
- Improved EMR on-cluster daemons to correctly track the node states when IP addresses are reused to improve reliability during scaling operations.
- SPARK-29683. Fixed an issue where job failures occurred during cluster scale-down as Spark was assuming all available nodes were deny-listed.

- [YARN-9011](). Fixed an issue where job failures occurred due to a race condition in YARN decommissioning when cluster tried to scale up or down.
- Fixed issue with step or job failures during cluster scaling by ensuring that the node states are always consistent between the Amazon EMR on-cluster daemons and YARN/HDFS.
- Fixed an issue where cluster operations such as scale down and step submission failed for Amazon EMR clusters enabled with Kerberos authentication. This was because the Amazon EMR on-cluster daemon did not renew the Kerberos ticket, which is required to securely communicate with HDFS/YARN running on the primary node.
- **Configuring a cluster to fix Apache YARN Timeline Server version 1 and 1.5 performance issues**

  Apache YARN Timeline Server version 1 and 1.5 can cause performance issues with very active, large EMR clusters, particularly with `yarn.resourcemanager.system-metrics-publisher.enabled=true`, which is the default setting in Amazon EMR. An open source YARN Timeline Server v2 solves the performance issue related to YARN Timeline Server scalability.

  Other workarounds for this issue include:

  - Configuring yarn.resourcemanager.system-metrics-publisher.enabled=false in yarn-site.xml.
  - Enabling the fix for this issue when creating a cluster, as described below.

  The following Amazon EMR releases contain a fix for this YARN Timeline Server performance issue.

  EMR 5.30.2, 5.31.1, 5.32.1, 5.33.1, 5.34.x, 6.0.1, 6.1.1, 6.2.1, 6.3.1, 6.4.x

  To enable the fix on any of the above specified Amazon EMR releases, set these properties to `true` in a configurations JSON file that is passed in using the [`aws emr create-cluster` command parameter](): `--configurations file://./configurations.json`. Or enable the fix using the [reconfiguration console UI]().

  Example of the configurations.json file contents:

```
[
{
"Classification": "yarn-site",
"Properties": {
"yarn.resourcemanager.system-metrics-publisher.timeline-server-v1.enable-batch":
 "true",
"yarn.resourcemanager.system-metrics-publisher.enabled": "true"
```

```
},
"Configurations": []
}
]
```

- WebHDFS and HttpFS server are disabled by default. You can re-enable WebHDFS using the Hadoop configuration, `dfs.webhdfs.enabled`. HttpFS server can be started by using `sudo systemctl start hadoop-httpfs`.
- HTTPS is now enabled by default for Amazon Linux repositories. If you are using an Amazon S3 VPCE policy to restrict access to specific buckets, you must add the new Amazon Linux bucket ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` to your policy (replace `$region` with the region where the endpoint is). For more information, see this topic in the AWS discussion forums. [Announcement: Amazon Linux 2 now supports the ability to use HTTPS while connecting to package repositories](#).
- Hive: Write query performance is improved by enabling the use of a scratch directory on HDFS for the last job. The temporary data for final job is written to HDFS instead of Amazon S3 and performance is improved because the data is moved from HDFS to the final table location (Amazon S3) instead of between Amazon S3 devices.
- Hive: Query compilation time improvement up to 2.5x with Glue metastore Partition Pruning.
- By default, when built-in UDFs are passed by Hive to the Hive Metastore Server, only a subset of those built-in UDFs are passed to the Glue Metastore since Glue supports only limited expression operators. If you set `hive.glue.partition.pruning.client=true`, then all partition pruning happens on the client side. If the you set `hive.glue.partition.pruning.server=true`, then all partition pruning happens on the server side.

Known issues

- Hue queries do not work in Amazon EMR 6.4.0 because Apache Hadoop HttpFS server is disabled by default. To use Hue on Amazon EMR 6.4.0, either manually start HttpFS server on the Amazon EMR primary node using `sudo systemctl start hadoop-httpfs`, or [use an Amazon EMR step](#).
- The Amazon EMR Notebooks feature used with Livy user impersonation does not work because HttpFS is disabled by default. In this case, the EMR notebook cannot connect to the cluster that has Livy impersonation enabled. The workaround is to start HttpFS server before connecting the EMR notebook to the cluster using `sudo systemctl start hadoop-httpfs`.

- In Amazon EMR version 6.4.0, Phoenix does not support the Phoenix connectors component.
- To use Spark actions with Apache Oozie, you must add the following configuration to your Oozie `workflow.xml` file. Otherwise, several critical libraries such as Hadoop and EMRFS will be missing from the classpath of the Spark executors that Oozie launches.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&‘()*+,-`. For more information, see [UTF-8 encoding table and Unicode characters](https://www.utf8-chartable.de/).

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

Release 5.32.0

The following release notes include information for Amazon EMR release 5.32.0. Changes are relative to 5.31.0.

Initial release date: Jan 8, 2021

Upgrades

- Upgraded Amazon Glue connector to version 1.14.0
- Upgraded Amazon SageMaker Spark SDK to version 1.4.1
- Upgraded AWS SDK for Java to version 1.11.890
- Upgraded EMR DynamoDB Connector version 4.16.0

- Upgraded EMRFS to version 2.45.0
- Upgraded EMR Log Analytics Metrics to version 1.18.0
- Upgraded EMR MetricsAndEventsApiGateway Client to version 1.5.0
- Upgraded EMR Record Server to version 1.8.0
- Upgraded EMR S3 Dist CP to version 2.17.0
- Upgraded EMR Secret Agent to version 1.7.0
- Upgraded Flink to version 1.11.2
- Upgraded Hadoop to version 2.10.1-amzn-0
- Upgraded Hive to version 2.3.7-amzn-3
- Upgraded Hue to version 4.8.0
- Upgraded Mxnet to version 1.7.0
- Upgraded OpenCV to version 4.4.0
- Upgraded Presto to version 0.240.1-amzn-0
- Upgraded Spark to version 2.4.7-amzn-0
- Upgraded TensorFlow to version 2.3.1

Changes, enhancements, and resolved issues

- This is a release to fix issues with Amazon EMR Scaling when it fails to scale up/scale down a cluster successfully or causes application failures.
- Fixed an issue where scaling requests failed for a large, highly utilized cluster when Amazon EMR on-cluster daemons were running health checking activities, such as gathering YARN node state and HDFS node state. This was happening because on-cluster daemons were not able to communicate the health status data of a node to internal Amazon EMR components.
- Improved EMR on-cluster daemons to correctly track the node states when IP addresses are reused to improve reliability during scaling operations.
- SPARK-29683. Fixed an issue where job failures occurred during cluster scale-down as Spark was assuming all available nodes were deny-listed.
- YARN-9011. Fixed an issue where job failures occurred due to a race condition in YARN decommissioning when cluster tried to scale up or down.
- Fixed issue with step or job failures during cluster scaling by ensuring that the node states are always consistent between the Amazon EMR on-cluster daemons and YARN/HDFS.

- Fixed an issue where cluster operations such as scale down and step submission failed for Amazon EMR clusters enabled with Kerberos authentication. This was because the Amazon EMR on-cluster daemon did not renew the Kerberos ticket, which is required to securely communicate with HDFS/YARN running on the primary node.
- Newer Amazon EMR releases fix the issue with a lower "Max open files" limit on older AL2 in Amazon EMR. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later now include a permanent fix with a higher "Max open files" setting.
- Upgraded component versions.
- For a list of component versions, see About Amazon EMR Releases in this guide.

New features

- Starting with Amazon EMR 5.32.0 and 6.5.0, dynamic executor sizing for Apache Spark is enabled by default. To turn this feature on or off, you can use the `spark.yarn.heterogeneousExecutors.enabled` configuration parameter.
- Instance Metadata Service (IMDS) V2 support status: Amazon EMR 5.23.1, 5.27.1 and 5.32 or later components use IMDSv2 for all IMDS calls. For IMDS calls in your application code, you can use both IMDSv1 and IMDSv2, or configure the IMDS to use only IMDSv2 for added security. For other 5.x EMR releases, disabling IMDSv1 causes cluster startup failure.
- Beginning with Amazon EMR 5.32.0, you can launch a cluster that natively integrates with Apache Ranger. Apache Ranger is an open-source framework to enable, monitor, and manage comprehensive data security across the Hadoop platform. For more information, see Apache Ranger. With native integration, you can bring your own Apache Ranger to enforce fine-grained data access control on Amazon EMR. See Integrate Amazon EMR with Apache Ranger in the *Amazon EMR Release Guide*.
- Amazon EMR Release 5.32.0 supports Amazon EMR on EKS. For more details on getting started with EMR on EKS, see What is Amazon EMR on EKS.
- Amazon EMR Release 5.32.0 supports Amazon EMR Studio (Preview). For more details on getting started with EMR Studio, see Amazon EMR Studio (Preview).
- Scoped managed policies: To align with AWS best practices, Amazon EMR has introduced v2 EMR-scoped default managed policies as replacements for policies that will be deprecated. See Amazon EMR Managed Policies.

Known issues

- For Amazon EMR 6.3.0 and 6.2.0 private subnet clusters, you cannot access the Ganglia web UI. You will get an "access denied (403)" error. Other web UIs, such as Spark, Hue, JupyterHub, Zeppelin, Livy, and Tez are working normally. Ganglia web UI access on public subnet clusters are also working normally. To resolve this issue, restart httpd service on the primary node with `sudo systemctl restart httpd`. This issue is fixed in Amazon EMR 6.4.0.
- **Lower "Max open files" limit on older AL2 [fixed in newer releases].** Amazon EMR releases: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0, and emr-6.2.0 are based on older versions ofAmazon Linux 2 (AL2), which have a lower ulimit setting for "Max open files" when Amazon EMR clusters are created with the default AMI. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later include a permanent fix with a higher "Max open files" setting. Releases with the lower open file limit causes a "Too many open files" error when submitting Spark job. In the impacted releases, the Amazon EMR default AMI has a default ulimit setting of 4096 for "Max open files," which is lower than the 65536 file limit in the latestAmazon Linux 2 AMI. The lower ulimit setting for "Max open files" causes Spark job failure when the Spark driver and executor try to open more than 4096 files. To fix the issue, Amazon EMR has a bootstrap action (BA) script that adjusts the ulimit setting at cluster creation.

  If you are using an older Amazon EMR version that doesn't have the permanent fix for this issue, the following workaround lets you to explicitly set the instance-controller ulimit to a maximum of 65536 files.

  Explicitly set a ulimit from the command line

  1. Edit `/etc/systemd/system/instance-controller.service` to add the following parameters to Service section.

     `LimitNOFILE=65536`

     `LimitNPROC=65536`
  2. Restart InstanceController

     `$ sudo systemctl daemon-reload`

     `$ sudo systemctl restart instance-controller`

  Set a ulimit using bootstrap action (BA)

You can also use a bootstrap action (BA) script to configure the instance-controller ulimit to 65536 files at cluster creation.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- **Important**

  EMR clusters that run Amazon Linux or Amazon Linux 2 Amazon Machine Images (AMIs) use default Amazon Linux behavior, and do not automatically download and install important and critical kernel updates that require a reboot. This is the same behavior as other Amazon EC2 instances that run the default Amazon Linux AMI. If new Amazon Linux software updates that require a reboot (such as kernel, NVIDIA, and CUDA updates) become available after an Amazon EMR release becomes available, EMR cluster instances that run the default AMI do not automatically download and install those updates. To get kernel updates, you can customize your Amazon EMR AMI to use the latest Amazon Linux AMI.

- Console support to create a security configuration that specifies the AWS Ranger integration option is currently not supported in the GovCloud Region. Security configuration can be done using the CLI. See Create the EMR Security Configuration in the *Amazon EMR Management Guide*.
- When AtRestEncryption or HDFS encryption is enabled on a cluster that uses Amazon EMR 5.31.0 or 5.32.0, Hive queries result in the following runtime exception.

```
TaskAttempt 3 failed, info=[Error: Error while running task ( failure ) :
 attempt_1604112648850_0001_1_01_000000_3:java.lang.RuntimeException:
 java.lang.RuntimeException: Hive Runtime Error while closing
 operators: java.io.IOException: java.util.ServiceConfigurationError:
 org.apache.hadoop.security.token.TokenIdentifier: Provider
 org.apache.hadoop.hbase.security.token.AuthenticationTokenIdentifier not found
```

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&‘()*+,-`. For more information, see [UTF-8 encoding table and Unicode characters](#).

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

Release 6.2.0

The following release notes include information for Amazon EMR release 6.2.0. Changes are relative to 6.1.0.

Initial release date: Dec 09, 2020

Last updated date: Oct 04, 2021

Supported applications

- AWS SDK for Java version 1.11.828
- emr-record-server version 1.7.0
- Flink version 1.11.2

- Ganglia version 3.7.2
- Hadoop version 3.2.1-amzn-1
- HBase version 2.2.6-amzn-0
- HBase-operator-tools 1.0.0
- HCatalog version 3.1.2-amzn-0
- Hive version 3.1.2-amzn-3
- Hudi version 0.6.0-amzn-1
- Hue version 4.8.0
- JupyterHub version 1.1.0
- Livy version 0.7.0
- MXNet version 1.7.0
- Oozie version 5.2.0
- Phoenix version 5.0.0
- Pig version 0.17.0
- Presto version 0.238.3-amzn-1
- PrestoSQL version 343
- Spark version 3.0.1-amzn-0
- spark-rapids 0.2.0
- TensorFlow version 2.3.1
- Zeppelin version 0.9.0-preview1
- Zookeeper version 3.4.14
- Connectors and drivers: DynamoDB Connector 4.16.0

New features

- HBase: Removed rename in commit phase and added persistent HFile tracking. See Persistent HFile Tracking in the *Amazon EMR Release Guide*.
- HBase: Backported Create a config that forces to cache blocks on compaction.
- PrestoDB: Improvements to Dynamic Partition Pruning. Rule-based Join Reorder works on non-partitioned data.

- Scoped managed policies: To align with AWS best practices, Amazon EMR has introduced v2 EMR-scoped default managed policies as replacements for policies that will be deprecated. See Amazon EMR Managed Policies.
- Instance Metadata Service (IMDS) V2 support status: For Amazon EMR 6.2 or later, Amazon EMR components use IMDSv2 for all IMDS calls. For IMDS calls in your application code, you can use both IMDSv1 and IMDSv2, or configure the IMDS to use only IMDSv2 for added security. If you disable IMDSv1 in earlier Amazon EMR 6.x releases, it causes cluster startup failure.

Changes, enhancements, and resolved issues

- This is a release to fix issues with Amazon EMR Scaling when it fails to scale up/scale down a cluster successfully or causes application failures.
- Fixed an issue where scaling requests failed for a large, highly utilized cluster when Amazon EMR on-cluster daemons were running health checking activities, such as gathering YARN node state and HDFS node state. This was happening because on-cluster daemons were not able to communicate the health status data of a node to internal Amazon EMR components.
- Improved EMR on-cluster daemons to correctly track the node states when IP addresses are reused to improve reliability during scaling operations.
- SPARK-29683. Fixed an issue where job failures occurred during cluster scale-down as Spark was assuming all available nodes were deny-listed.
- YARN-9011. Fixed an issue where job failures occurred due to a race condition in YARN decommissioning when cluster tried to scale up or down.
- Fixed issue with step or job failures during cluster scaling by ensuring that the node states are always consistent between the Amazon EMR on-cluster daemons and YARN/HDFS.
- Fixed an issue where cluster operations such as scale down and step submission failed for Amazon EMR clusters enabled with Kerberos authentication. This was because the Amazon EMR on-cluster daemon did not renew the Kerberos ticket, which is required to securely communicate with HDFS/YARN running on the primary node.
- Newer Amazon EMR releases fix the issue with a lower "Max open files" limit on older AL2 in Amazon EMR. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later now include a permanent fix with a higher "Max open files" setting.
- Spark: Performance improvements in Spark runtime.

Known issues

- Amazon EMR 6.2 has incorrect permissions set on the /etc/cron.d/libinstance-controller-java file in EMR 6.2.0. Permissions on the file are 645 (-rw-r--r-x), when they should be 644 (-rw-r--r--). As a result, Amazon EMR version 6.2 does not log instance-state logs, and the /emr/instance-logs directory is empty. This issue is fixed in Amazon EMR 6.3.0 and later.

  To work around this issue, run the following script as a bootstrap action at cluster launch.

  ```
  #!/bin/bash
  sudo chmod 644 /etc/cron.d/libinstance-controller-java
  ```

- For Amazon EMR 6.2.0 and 6.3.0 private subnet clusters, you cannot access the Ganglia web UI. You will get an "access denied (403)" error. Other web UIs, such as Spark, Hue, JupyterHub, Zeppelin, Livy, and Tez are working normally. Ganglia web UI access on public subnet clusters are also working normally. To resolve this issue, restart httpd service on the primary node with `sudo systemctl restart httpd`. This issue is fixed in Amazon EMR 6.4.0.
- There is an issue in Amazon EMR 6.2.0 where httpd continuously fails, causing Ganglia to be unavailable. You get a "cannot connect to the server" error. To fix a cluster that is already running with this issue, SSH to the cluster primary node and add the line `Listen 80` to the file `httpd.conf` located at `/etc/httpd/conf/httpd.conf`. This issue is fixed in Amazon EMR 6.3.0.
- HTTPD fails on EMR 6.2.0 clusters when you use a security configuration. This makes the Ganglia web application user interface unavailable. To access the Ganglia web application user interface, add `Listen 80` to the `/etc/httpd/conf/httpd.conf` file on the primary node of your cluster. For information about connecting to your cluster, see Connect to the Primary Node Using SSH.

  EMR Notebooks also fail to establish a connection with EMR 6.2.0 clusters when you use a security configuration. The notebook will fail to list kernels and submit Spark jobs. We recommend that you use EMR Notebooks with another version of Amazon EMR instead.
- **Lower "Max open files" limit on older AL2 [fixed in newer releases].** Amazon EMR releases: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0, and emr-6.2.0 are based on older versions ofAmazon Linux 2 (AL2), which have a lower ulimit setting for "Max open files" when Amazon EMR clusters are created with the default AMI. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later include a permanent fix with a higher "Max open files" setting. Releases with the lower open file limit causes a "Too many open files"

error when submitting Spark job. In the impacted releases, the Amazon EMR default AMI has a default ulimit setting of 4096 for "Max open files," which is lower than the 65536 file limit in the latestAmazon Linux 2 AMI. The lower ulimit setting for "Max open files" causes Spark job failure when the Spark driver and executor try to open more than 4096 files. To fix the issue, Amazon EMR has a bootstrap action (BA) script that adjusts the ulimit setting at cluster creation.

If you are using an older Amazon EMR version that doesn't have the permanent fix for this issue, the following workaround lets you to explicitly set the instance-controller ulimit to a maximum of 65536 files.

Explicitly set a ulimit from the command line

1. Edit `/etc/systemd/system/instance-controller.service` to add the following parameters to Service section.

   `LimitNOFILE=65536`

   `LimitNPROC=65536`
2. Restart InstanceController

   `$ sudo systemctl daemon-reload`

   `$ sudo systemctl restart instance-controller`

Set a ulimit using bootstrap action (BA)

You can also use a bootstrap action (BA) script to configure the instance-controller ulimit to 65536 files at cluster creation.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
```

```
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- **Important**

  Amazon EMR 6.1.0 and 6.2.0 include a performance issue that can critically affect all Hudi insert, upsert, and delete operations. If you plan to use Hudi with Amazon EMR 6.1.0 or 6.2.0, you should contact AWS support to obtain a patched Hudi RPM.

- **Important**

  EMR clusters that run Amazon Linux or Amazon Linux 2 Amazon Machine Images (AMIs) use default Amazon Linux behavior, and do not automatically download and install important and critical kernel updates that require a reboot. This is the same behavior as other Amazon EC2 instances that run the default Amazon Linux AMI. If new Amazon Linux software updates that require a reboot (such as kernel, NVIDIA, and CUDA updates) become available after an Amazon EMR release becomes available, EMR cluster instances that run the default AMI do not automatically download and install those updates. To get kernel updates, you can customize your Amazon EMR AMI to use the latest Amazon Linux AMI.

- Amazon EMR 6.2.0 Maven artifacts are not published. They will be published with a future release of Amazon EMR.
- Persistent HFile tracking using the HBase storefile system table does not support the HBase region replication feature. For more information about HBase region replication, see Timeline-consistent High Available Reads.
- Amazon EMR 6.x and EMR 5.x Hive bucketing version differences

  EMR 5.x uses OOS Apache Hive 2, while in EMR 6.x uses OOS Apache Hive 3. The open source Hive2 uses Bucketing version 1, while open source Hive3 uses Bucketing version 2. This bucketing version difference between Hive 2 (EMR 5.x) and Hive 3 (EMR 6.x) means Hive bucketing hashing functions differently. See the example below.

  The following table is an example created in EMR 6.x and EMR 5.x, respectively.

```
-- Using following LOCATION in EMR 6.x
CREATE TABLE test_bucketing (id INT, desc STRING)
PARTITIONED BY (day STRING)
CLUSTERED BY(id) INTO 128 BUCKETS
LOCATION 's3://your-own-s3-bucket/emr-6-bucketing/';

-- Using following LOCATION in EMR 5.x
LOCATION 's3://your-own-s3-bucket/emr-5-bucketing/';
```

Inserting the same data in both EMR 6.x and EMR 5.x.

```
INSERT INTO test_bucketing PARTITION (day='01') VALUES(66, 'some_data');
INSERT INTO test_bucketing PARTITION (day='01') VALUES(200, 'some_data');
```

Checking the S3 location, shows the bucketing file name is different, because the hashing function is different between EMR 6.x (Hive 3) and EMR 5.x (Hive 2).

```
[hadoop@ip-10-0-0-122 ~]$ aws s3 ls s3://your-own-s3-bucket/emr-6-bucketing/day=01/
2020-10-21 20:35:16         13 000025_0
2020-10-21 20:35:22         14 000121_0
[hadoop@ip-10-0-0-122 ~]$ aws s3 ls s3://your-own-s3-bucket/emr-5-bucketing/day=01/
2020-10-21 20:32:07         13 000066_0
2020-10-21 20:32:51         14 000072_0
```

You can also see the version difference by running the following command in Hive CLI in EMR 6.x. Note that it returns bucketing version 2.

```
hive> DESCRIBE FORMATTED test_bucketing;
...
Table Parameters:
    bucketing_version       2
...
```

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem

impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

Workaround:

- SSH as hadoop user to the lead primary node of the EMR cluster with multiple primary nodes.
- Run the following command to renew Kerberos ticket for hadoop user.

```
kinit -kt <keytab_file> <principal>
```

Typically, the keytab file is located at /etc/hadoop.keytab and the principal is in the form of hadoop/<hostname>@<REALM>.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, s3://bucket/table/p=a is a prefix of s3://bucket/table/p=a b.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in s3://bucket/table/p=a b falls into this category. Note that there are 14 other non-control characters: !"#$%&‘()*+,-. For more information, see UTF-8 encoding table and Unicode characters.

  As a workaround to this issue, set the spark.sql.sources.fastS3PartitionDiscovery.enabled configuration to false in the spark-defaults classification.

Release 5.31.0

The following release notes include information for Amazon EMR release 5.31.0. Changes are relative to 5.30.1.

Initial release date: Oct 9, 2020

Last updated date: Oct 15, 2020

Upgrades

- Upgraded Amazon Glue connector to version 1.13.0
- Upgraded Amazon SageMaker Spark SDK to version 1.4.0
- Upgraded Amazon Kinesis connector to version 3.5.9
- Upgraded AWS SDK for Java to version 1.11.852
- Upgraded Bigtop-tomcat to version 8.5.56
- Upgraded EMR FS to version 2.43.0
- Upgraded EMR MetricsAndEventsApiGateway Client to version 1.4.0
- Upgraded EMR S3 Dist CP to version 2.15.0
- Upgraded EMR S3 Select to version 1.6.0
- Upgraded Flink to version 1.11.0
- Upgraded Hadoop to version 2.10.0
- Upgraded Hive to version 2.3.7
- Upgraded Hudi to version 0.6.0
- Upgraded Hue to version 4.7.1
- Upgraded JupyterHub to version 1.1.0
- Upgraded Mxnet to version 1.6.0
- Upgraded OpenCV to version 4.3.0
- Upgraded Presto to version 0.238.3
- Upgraded TensorFlow to version 2.1.0

Changes, enhancements, and resolved issues

- This is a release to fix issues with Amazon EMR Scaling when it fails to scale up/scale down a cluster successfully or causes application failures.

- Fixed an issue where scaling requests failed for a large, highly utilized cluster when Amazon EMR on-cluster daemons were running health checking activities, such as gathering YARN node state and HDFS node state. This was happening because on-cluster daemons were not able to communicate the health status data of a node to internal Amazon EMR components.
- Improved EMR on-cluster daemons to correctly track the node states when IP addresses are reused to improve reliability during scaling operations.
- SPARK-29683. Fixed an issue where job failures occurred during cluster scale-down as Spark was assuming all available nodes were deny-listed.
- YARN-9011. Fixed an issue where job failures occurred due to a race condition in YARN decommissioning when cluster tried to scale up or down.
- Fixed issue with step or job failures during cluster scaling by ensuring that the node states are always consistent between the Amazon EMR on-cluster daemons and YARN/HDFS.
- Fixed an issue where cluster operations such as scale down and step submission failed for Amazon EMR clusters enabled with Kerberos authentication. This was because the Amazon EMR on-cluster daemon did not renew the Kerberos ticket, which is required to securely communicate with HDFS/YARN running on the primary node.
- Newer Amazon EMR releases fix the issue with a lower "Max open files" limit on older AL2 in Amazon EMR. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later now include a permanent fix with a higher "Max open files" setting.
- Hive column statistics are supported for Amazon EMR versions 5.31.0 and later.
- Upgraded component versions.
- EMRFS S3EC V2 Support in Amazon EMR 5.31.0. In S3 Java SDK releases 1.11.837 and later, encryption client Version 2 (S3EC V2) has been introduced with various security enhancements. For more information, see the following:
  - S3 blog post: Updates to the Amazon S3 encryption client.
  - AWS SDK for Java Developer Guide: Migrate encryption and decryption clients to V2.
  - EMR Management Guide: Amazon S3 client-side encryption.

  Encryption Client V1 is still available in the SDK for backward compatibility.

New features

- **Lower "Max open files" limit on older AL2 [fixed in newer releases].** Amazon EMR releases: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0, and emr-6.2.0 are based on older versions ofAmazon Linux 2 (AL2), which have a lower ulimit setting for "Max open files" when Amazon EMR clusters are created with the default AMI. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later include a permanent fix with a higher "Max open files" setting. Releases with the lower open file limit causes a "Too many open files" error when submitting Spark job. In the impacted releases, the Amazon EMR default AMI has a default ulimit setting of 4096 for "Max open files," which is lower than the 65536 file limit in the latestAmazon Linux 2 AMI. The lower ulimit setting for "Max open files" causes Spark job failure when the Spark driver and executor try to open more than 4096 files. To fix the issue, Amazon EMR has a bootstrap action (BA) script that adjusts the ulimit setting at cluster creation.

If you are using an older Amazon EMR version that doesn't have the permanent fix for this issue, the following workaround lets you to explicitly set the instance-controller ulimit to a maximum of 65536 files.

Explicitly set a ulimit from the command line

1. Edit `/etc/systemd/system/instance-controller.service` to add the following parameters to Service section.

   `LimitNOFILE=65536`

   `LimitNPROC=65536`
2. Restart InstanceController

   `$ sudo systemctl daemon-reload`

   `$ sudo systemctl restart instance-controller`

Set a ulimit using bootstrap action (BA)

You can also use a bootstrap action (BA) script to configure the instance-controller ulimit to 65536 files at cluster creation.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
```

```
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- With Amazon EMR 5.31.0, you can launch a cluster that integrates with Lake Formation. This integration provides fine-grained, column-level data filtering to databases and tables in the AWS Glue Data Catalog. It also enables federated single sign-on to EMR Notebooks or Apache Zeppelin from an enterprise identity system. For more information, see Integrating Amazon EMR with AWS Lake Formation in the *Amazon EMR Management Guide*.

  Amazon EMR with Lake Formation is currently available in 16 AWS Regions: US East (Ohio and N. Virginia), US West (N. California and Oregon), Asia Pacific (Mumbai, Seoul, Singapore, Sydney, and Tokyo), Canada (Central), Europe (Frankfurt, Ireland, London, Paris, and Stockholm), South America (São Paulo).

Known issues

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as hadoop user to the lead primary node of the EMR cluster with multiple primary nodes.
  - Run the following command to renew Kerberos ticket for hadoop user.

  ```
  kinit -kt <keytab_file> <principal>
  ```

Typically, the keytab file is located at `/etc/hadoop.keytab` and the principal is in the form of `hadoop/<hostname>@<REALM>`.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

- When AtRestEncryption or HDFS encryption is enabled on a cluster that uses Amazon EMR 5.31.0 or 5.32.0, Hive queries result in the following runtime exception.

```
TaskAttempt 3 failed, info=[Error: Error while running task ( failure ) :
 attempt_1604112648850_0001_1_01_000000_3:java.lang.RuntimeException:
 java.lang.RuntimeException: Hive Runtime Error while closing
 operators: java.io.IOException: java.util.ServiceConfigurationError:
 org.apache.hadoop.security.token.TokenIdentifier: Provider
 org.apache.hadoop.hbase.security.token.AuthenticationTokenIdentifier not found
```

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&'()*+,-`. For more information, see [UTF-8 encoding table and Unicode characters](https://www.utf8-chartable.de/).

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

Release 6.1.0

The following release notes include information for Amazon EMR release 6.1.0. Changes are relative to 6.0.0.

Initial release date: Sept 04, 2020

Last updated date: Oct 15, 2020

Supported applications

- AWS SDK for Java version 1.11.828
- Flink version 1.11.0
- Ganglia version 3.7.2
- Hadoop version 3.2.1-amzn-1
- HBase version 2.2.5
- HBase-operator-tools 1.0.0
- HCatalog version 3.1.2-amzn-0
- Hive version 3.1.2-amzn-1
- Hudi version 0.5.2-incubating
- Hue version 4.7.1
- JupyterHub version 1.1.0
- Livy version 0.7.0
- MXNet version 1.6.0
- Oozie version 5.2.0
- Phoenix version 5.0.0
- Presto version 0.232
- PrestoSQL version 338
- Spark version 3.0.0-amzn-0
- TensorFlow version 2.1.0
- Zeppelin version 0.9.0-preview1
- Zookeeper version 3.4.14

- Connectors and drivers: DynamoDB Connector 4.14.0

New features

- ARM instance types are supported starting with Amazon EMR version 5.30.0 and Amazon EMR version 6.1.0.
- M6g general purpose instance types are supported starting with Amazon EMR versions 6.1.0 and 5.30.0. For more information, see Supported Instance Types in the *Amazon EMR Management Guide*.
- The EC2 placement group feature is supported starting with Amazon EMR version 5.23.0 as an option for multiple primary node clusters. Currently, only primary node types are supported by the placement group feature, and the `SPREAD` strategy is applied to those primary nodes. The `SPREAD` strategy places a small group of instances across separate underlying hardware to guard against the loss of multiple primary nodes in the event of a hardware failure. For more information, see EMR Integration with EC2 Placement Group in the *Amazon EMR Management Guide*.
- Managed Scaling – With Amazon EMR version 6.1.0, you can enable Amazon EMR managed scaling to automatically increase or decrease the number of instances or units in your cluster based on workload. Amazon EMR continuously evaluates cluster metrics to make scaling decisions that optimize your clusters for cost and speed. Managed Scaling is also available on Amazon EMR version 5.30.0 and later, except 6.0.0. For more information, see Scaling Cluster Resources in the *Amazon EMR Management Guide*.
- PrestoSQL version 338 is supported with EMR 6.1.0. For more information, see Presto.
  - PrestoSQL is supported on EMR 6.1.0 and later versions only, not on EMR 6.0.0 or EMR 5.x.
  - The application name, `Presto` continues to be used to install PrestoDB on clusters. To install PrestoSQL on clusters, use the application name `PrestoSQL`.
  - You can install either PrestoDB or PrestoSQL, but you cannot install both on a single cluster. If both PrestoDB and PrestoSQL are specified when attempting to create a cluster, a validation error occurs and the cluster creation request fails.
  - PrestoSQL is supported on both single-master and muti-master clusters. On multi-master clusters, an external Hive metastore is required to run PrestoSQL or PrestoDB. See Supported applications in an EMR cluster with multiple primary nodes.
- ECR auto authentication support on Apache Hadoop and Apache Spark with Docker: Spark users can use Docker images from Docker Hub and Amazon Elastic Container Registry (Amazon ECR) to define environment and library dependencies.

Configure Docker and Run Spark Applications with Docker Using Amazon EMR 6.x.

- EMR supports Apache Hive ACID transactions: Amazon EMR 6.1.0 adds support for Hive ACID transactions so it complies with the ACID properties of a database. With this feature, you can run `INSERT, UPDATE, DELETE,` and `MERGE` operations in Hive managed tables with data in Amazon Simple Storage Service (Amazon S3). This is a key feature for use cases like streaming ingestion, data restatement, bulk updates using MERGE, and slowly changing dimensions. For more information, including configuration examples and use cases, see Amazon EMR supports Apache Hive ACID transactions.

Changes, enhancements, and resolved issues

- This is a release to fix issues with Amazon EMR Scaling when it fails to scale up/scale down a cluster successfully or causes application failures.
- Fixed an issue where scaling requests failed for a large, highly utilized cluster when Amazon EMR on-cluster daemons were running health checking activities, such as gathering YARN node state and HDFS node state. This was happening because on-cluster daemons were not able to communicate the health status data of a node to internal Amazon EMR components.
- Improved EMR on-cluster daemons to correctly track the node states when IP addresses are reused to improve reliability during scaling operations.
- SPARK-29683. Fixed an issue where job failures occurred during cluster scale-down as Spark was assuming all available nodes were deny-listed.
- YARN-9011. Fixed an issue where job failures occurred due to a race condition in YARN decommissioning when cluster tried to scale up or down.
- Fixed issue with step or job failures during cluster scaling by ensuring that the node states are always consistent between the Amazon EMR on-cluster daemons and YARN/HDFS.
- Fixed an issue where cluster operations such as scale down and step submission failed for Amazon EMR clusters enabled with Kerberos authentication. This was because the Amazon EMR on-cluster daemon did not renew the Kerberos ticket, which is required to securely communicate with HDFS/YARN running on the primary node.
- Newer Amazon EMR releases fix the issue with a lower "Max open files" limit on older AL2 in Amazon EMR. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later now include a permanent fix with a higher "Max open files" setting.

- Apache Flink is not supported on EMR 6.0.0, but it is supported on EMR 6.1.0 with Flink 1.11.0. This is the first version of Flink to officially support Hadoop 3. See Apache Flink 1.11.0 Release Announcement.
- Ganglia has been removed from default EMR 6.1.0 package bundles.

Known issues

- **Lower "Max open files" limit on older AL2 [fixed in newer releases].** Amazon EMR releases: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0, and emr-6.2.0 are based on older versions ofAmazon Linux 2 (AL2), which have a lower ulimit setting for "Max open files" when Amazon EMR clusters are created with the default AMI. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later include a permanent fix with a higher "Max open files" setting. Releases with the lower open file limit causes a "Too many open files" error when submitting Spark job. In the impacted releases, the Amazon EMR default AMI has a default ulimit setting of 4096 for "Max open files," which is lower than the 65536 file limit in the latestAmazon Linux 2 AMI. The lower ulimit setting for "Max open files" causes Spark job failure when the Spark driver and executor try to open more than 4096 files. To fix the issue, Amazon EMR has a bootstrap action (BA) script that adjusts the ulimit setting at cluster creation.

  If you are using an older Amazon EMR version that doesn't have the permanent fix for this issue, the following workaround lets you to explicitly set the instance-controller ulimit to a maximum of 65536 files.

  Explicitly set a ulimit from the command line

  1. Edit `/etc/systemd/system/instance-controller.service` to add the following parameters to Service section.

     ```
     LimitNOFILE=65536
     ```

     ```
     LimitNPROC=65536
     ```
  2. Restart InstanceController

     ```
     $ sudo systemctl daemon-reload
     ```

     ```
     $ sudo systemctl restart instance-controller
     ```

  Set a ulimit using bootstrap action (BA)

You can also use a bootstrap action (BA) script to configure the instance-controller ulimit to 65536 files at cluster creation.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- **Important**

  Amazon EMR 6.1.0 and 6.2.0 include a performance issue that can critically affect all Hudi insert, upsert, and delete operations. If you plan to use Hudi with Amazon EMR 6.1.0 or 6.2.0, you should contact AWS support to obtain a patched Hudi RPM.

- If you set custom garbage collection configuration with `spark.driver.extraJavaOptions` and `spark.executor.extraJavaOptions`, this will result in driver/executor launch failure with EMR 6.1 due to conflicting garbage collection configuration. With EMR Release 6.1.0, you should specify custom Spark garbage collection configuration for drivers and executors with the properties `spark.driver.defaultJavaOptions` and `spark.executor.defaultJavaOptions` instead. Read more in Apache Spark Runtime Environment and Configuring Spark Garbage Collection on Amazon EMR 6.1.0.
- Using Pig with Oozie (and within Hue, since Hue uses Oozie actions to run Pig scripts), generates an error that a native-lzo library cannot be loaded. This error message is informational and does not block Pig from running.

- Hudi Concurrency Support: Currently Hudi doesn't support concurrent writes to a single Hudi table. In addition, Hudi rolls back any changes being done by in-progress writers before allowing a new writer to start. Concurrent writes can interfere with this mechanism and introduce race conditions, which can lead to data corruption. You should ensure that as part of your data processing workflow, there is only a single Hudi writer operating against a Hudi table at any time. Hudi does support multiple concurrent readers operating against the same Hudi table.
- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as hadoop user to the lead primary node of the EMR cluster with multiple primary nodes.
  - Run the following command to renew Kerberos ticket for hadoop user.

    ```
    kinit -kt <keytab_file> <principal>
    ```

    Typically, the keytab file is located at /etc/hadoop.keytab and the principal is in the form of hadoop/<hostname>@<REALM>.

  Note

  This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

- There is an issue in Amazon EMR 6.1.0 that affects clusters running Presto. After an extended period of time (days), the cluster may throw errors such as, "su: failed to execute /bin/bash: Resource temporarily unavailable" or "shell request failed on channel 0". This issue is caused by an internal Amazon EMR process (InstanceController) that is spawning too many Light Weight Processes (LWP), which eventually causes the Hadoop user to exceed their nproc limit. This prevents the user from opening additional processes. The solution for this issue is to upgrade to EMR 6.2.0.

Release 6.0.0

The following release notes include information for Amazon EMR release 6.0.0.

Initial release date: March 10, 2020

Supported applications

- AWS SDK for Java version 1.11.711
- Ganglia version 3.7.2
- Hadoop version 3.2.1
- HBase version 2.2.3
- HCatalog version 3.1.2
- Hive version 3.1.2
- Hudi version 0.5.0-incubating
- Hue version 4.4.0
- JupyterHub version 1.0.0
- Livy version 0.6.0
- MXNet version 1.5.1
- Oozie version 5.1.0
- Phoenix version 5.0.0
- Presto version 0.230
- Spark version 2.4.4
- TensorFlow version 1.14.0
- Zeppelin version 0.9.0-SNAPSHOT
- Zookeeper version 3.4.14
- Connectors and drivers: DynamoDB Connector 4.14.0

> **Note**
>
> Flink, Sqoop, Pig, and Mahout are not available in Amazon EMR version 6.0.0.

New features

- YARN Docker Runtime Support - YARN applications, such as Spark jobs, can now run in the context of a Docker container. This allows you to easily define dependencies in a Docker image without the need to install custom libraries on your Amazon EMR cluster. For more information, see Configure Docker Integration and Run Spark applications with Docker using Amazon EMR 6.0.0.
- Hive LLAP Support - Hive now supports the LLAP execution mode for improved query performance. For more information, see Using Hive LLAP.

Changes, enhancements, and resolved issues

- This is a release to fix issues with Amazon EMR Scaling when it fails to scale up/scale down a cluster successfully or causes application failures.
- Fixed an issue where scaling requests failed for a large, highly utilized cluster when Amazon EMR on-cluster daemons were running health checking activities, such as gathering YARN node state and HDFS node state. This was happening because on-cluster daemons were not able to communicate the health status data of a node to internal Amazon EMR components.
- Improved EMR on-cluster daemons to correctly track the node states when IP addresses are reused to improve reliability during scaling operations.
- SPARK-29683. Fixed an issue where job failures occurred during cluster scale-down as Spark was assuming all available nodes were deny-listed.
- YARN-9011. Fixed an issue where job failures occurred due to a race condition in YARN decommissioning when cluster tried to scale up or down.
- Fixed issue with step or job failures during cluster scaling by ensuring that the node states are always consistent between the Amazon EMR on-cluster daemons and YARN/HDFS.
- Fixed an issue where cluster operations such as scale down and step submission failed for Amazon EMR clusters enabled with Kerberos authentication. This was because the Amazon EMR on-cluster daemon did not renew the Kerberos ticket, which is required to securely communicate with HDFS/YARN running on the primary node.
- Newer Amazon EMR releases fix the issue with a lower "Max open files" limit on older AL2 in Amazon EMR. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later now include a permanent fix with a higher "Max open files" setting.
- Amazon Linux
  - Amazon Linux 2 is the operating system for the EMR 6.x release series.

  - `systemd` is used for service management instead of `upstart` used inAmazon Linux 1.
- Java Development Kit (JDK)
  - Corretto JDK 8 is the default JDK for the EMR 6.x release series.
- Scala
  - Scala 2.12 is used with Apache Spark and Apache Livy.
- Python 3
  - Python 3 is now the default version of Python in EMR.
- YARN node labels
  - Beginning with Amazon EMR 6.x release series, the YARN node labels feature is disabled by default. The application master processes can run on both core and task nodes by default. You can enable the YARN node labels feature by configuring following properties: `yarn.node-labels.enabled` and `yarn.node-labels.am.default-node-label-expression`. For more information, see Understanding Primary, Core, and Task Nodes.

Known issues

- **Lower "Max open files" limit on older AL2 [fixed in newer releases].** Amazon EMR releases: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0, and emr-6.2.0 are based on older versions ofAmazon Linux 2 (AL2), which have a lower ulimit setting for "Max open files" when Amazon EMR clusters are created with the default AMI. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later include a permanent fix with a higher "Max open files" setting. Releases with the lower open file limit causes a "Too many open files" error when submitting Spark job. In the impacted releases, the Amazon EMR default AMI has a default ulimit setting of 4096 for "Max open files," which is lower than the 65536 file limit in the latestAmazon Linux 2 AMI. The lower ulimit setting for "Max open files" causes Spark job failure when the Spark driver and executor try to open more than 4096 files. To fix the issue, Amazon EMR has a bootstrap action (BA) script that adjusts the ulimit setting at cluster creation.

  If you are using an older Amazon EMR version that doesn't have the permanent fix for this issue, the following workaround lets you to explicitly set the instance-controller ulimit to a maximum of 65536 files.

  Explicitly set a ulimit from the command line

  1. Edit `/etc/systemd/system/instance-controller.service` to add the following parameters to Service section.

```
LimitNOFILE=65536

LimitNPROC=65536
```

2. Restart InstanceController

```
$ sudo systemctl daemon-reload

$ sudo systemctl restart instance-controller
```

Set a ulimit using bootstrap action (BA)

You can also use a bootstrap action (BA) script to configure the instance-controller ulimit to 65536 files at cluster creation.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- Spark interactive shell, including PySpark, SparkR, and spark-shell, does not support using Docker with additional libraries.
- To use Python 3 with Amazon EMR version 6.0.0, you must add PATH to `yarn.nodemanager.env-whitelist`.
- The Live Long and Process (LLAP) functionality is not supported when you use the AWS Glue Data Catalog as the metastore for Hive.

- When using Amazon EMR 6.0.0 with Spark and Docker integration, you need to configure the instances in your cluster with the same instance type and the same amount of EBS volumes to avoid failure when submitting a Spark job with Docker runtime.
- In Amazon EMR 6.0.0, HBase on Amazon S3 storage mode is impacted by the HBASE-24286. issue. HBase master cannot initialize when the cluster is created using existing S3 data.
- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as hadoop user to the lead primary node of the EMR cluster with multiple primary nodes.
  - Run the following command to renew Kerberos ticket for hadoop user.

    ```
    kinit -kt <keytab_file> <principal>
    ```

    Typically, the keytab file is located at /etc/hadoop.keytab and the principal is in the form of hadoop/<hostname>@<REALM>.

  Note

  This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

Release 5.30.1

The following release notes include information for Amazon EMR release 5.30.1. Changes are relative to 5.30.0.

Initial release date: June 30, 2020

Last updated date: August 24, 2020

Changes, enhancements, and resolved issues

- Newer Amazon EMR releases fix the issue with a lower "Max open files" limit on older AL2 in Amazon EMR. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later now include a permanent fix with a higher "Max open files" setting.
- Fixed issue where instance controller process spawned infinite number of processes.
- Fixed issue where Hue was unable to run an Hive query, showing a "database is locked" message and preventing the execution of queries.
- Fixed a Spark issue to enable more tasks to run concurrently on the EMR cluster.
- Fixed a Jupyter notebook issue causing a "too many files open error" in the Jupyter server.
- Fixed an issue with cluster start times.

New features

- Tez UI and YARN timeline server persistent application interfaces are available with Amazon EMR versions 6.x, and EMR version 5.30.1 and later. One-click link access to persistent application history lets you quickly access job history without setting up a web proxy through an SSH connection. Logs for active and terminated clusters are available for 30 days after the application ends. For more information, see View Persistent Application User Interfaces in the *Amazon EMR Management Guide*.
- EMR Notebook execution APIs are available to execute EMR notebooks via a script or command line. The ability to start, stop, list, and describe EMR notebook executions without the AWS console enables you programmatically control an EMR notebook. Using a parameterized notebook cell, you can pass different parameter values to a notebook without having to create a copy of the notebook for each new set of paramter values. See EMR API Actions. For sample code, see Sample commands to execute EMR Notebooks programmatically.

Known issues

- **Lower "Max open files" limit on older AL2 [fixed in newer releases].** Amazon EMR releases: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0, and emr-6.2.0 are based on older versions ofAmazon Linux 2 (AL2), which have a lower ulimit setting for "Max open files" when Amazon EMR clusters are created with the default AMI. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later include a permanent fix with a higher "Max open files" setting. Releases with the lower open file limit causes a "Too many open files" error when submitting Spark job. In the impacted releases, the Amazon EMR default AMI has a default ulimit setting of 4096 for "Max open files," which is lower than the 65536 file limit in the latestAmazon Linux 2 AMI. The lower ulimit setting for "Max open files" causes Spark job failure when the Spark driver and executor try to open more than 4096 files. To fix the issue, Amazon EMR has a bootstrap action (BA) script that adjusts the ulimit setting at cluster creation.

If you are using an older Amazon EMR version that doesn't have the permanent fix for this issue, the following workaround lets you to explicitly set the instance-controller ulimit to a maximum of 65536 files.

Explicitly set a ulimit from the command line

1. Edit `/etc/systemd/system/instance-controller.service` to add the following parameters to Service section.

   `LimitNOFILE=65536`

   `LimitNPROC=65536`

2. Restart InstanceController

   `$ sudo systemctl daemon-reload`

   `$ sudo systemctl restart instance-controller`

Set a ulimit using bootstrap action (BA)

You can also use a bootstrap action (BA) script to configure the instance-controller ulimit to 65536 files at cluster creation.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
```

```
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- **EMR Notebooks**

  The feature that allows you to install kernels and additional Python libraries on the cluster primary node is disabled by default on EMR version 5.30.1. For more information about this feature, see Installing Kernels and Python Libraries on a Cluster Primary Node.

  To enable the feature, do the following:

  1. Make sure that the permissions policy attached to the service role for EMR Notebooks allows the following action:

     `elasticmapreduce:ListSteps`

     For more information, see Service Role for EMR Notebooks.
  2. Use the AWS CLI to run a step on the cluster that sets up EMR Notebooks as shown in the following example. Replace *us-east-1* with the Region in which your cluster resides. For more information, see Adding Steps to a Cluster Using the AWS CLI.

     ```
     aws emr add-steps  --cluster-id MyClusterID --steps
      Type=CUSTOM_JAR,Name=EMRNotebooksSetup,ActionOnFailure=CONTINUE,Jar=s3://us-
     east-1.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://
     awssupportdatasvcs.com/bootstrap-actions/EMRNotebooksSetup/emr-notebooks-setup.sh"]
     ```

- **Managed scaling**

  Managed scaling operations on 5.30.0 and 5.30.1 clusters without Presto installed may cause application failures or cause a uniform instance group or instance fleet to stay in the ARRESTED state, particularly when a scale down operation is followed quickly by a scale up operation.

  As a workaround, choose Presto as an application to install when you create a cluster with Amazon EMR releases 5.30.0 and 5.30.1, even if your job does not require Presto.
- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

Workaround:

- SSH as `hadoop` user to the lead primary node of the EMR cluster with multiple primary nodes.
- Run the following command to renew Kerberos ticket for `hadoop` user.

```
kinit -kt <keytab_file> <principal>
```

  Typically, the keytab file is located at `/etc/hadoop.keytab` and the principal is in the form of `hadoop/<hostname>@<REALM>`.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&‘()*+,-`. For more information, see UTF-8 encoding table and Unicode characters.

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

Release 5.30.0

The following release notes include information for Amazon EMR release 5.30.0. Changes are relative to 5.29.0.

Initial release date: May 13, 2020

Last updated date: June 25, 2020

Upgrades

- Upgraded AWS SDK for Java to version 1.11.759
- Upgraded Amazon SageMaker Spark SDK to version 1.3.0
- Upgraded EMR Record Server to version 1.6.0
- Upgraded Flink to version 1.10.0
- Upgraded Ganglia to version 3.7.2
- Upgraded HBase to version 1.4.13
- Upgraded Hudi to version 0.5.2-incubating
- Upgraded Hue to version 4.6.0
- Upgraded JupyterHub to version 1.1.0
- Upgraded Livy to version 0.7.0-incubating
- Upgraded Oozie to version 5.2.0
- Upgraded Presto to version 0.232
- Upgraded Spark to version 2.4.5
- Upgraded Connectors and drivers: Amazon Glue Connector 1.12.0; Amazon Kinesis Connector 3.5.0; EMR DynamoDB Connector 4.14.0

New features

- **EMR Notebooks** – When used with EMR clusters created using 5.30.0, EMR notebook kernels run on cluster. This improves notebook performance and allows you to install and customize kernels. You can also install Python libraries on the cluster primary node. For more information, see Installing and Using Kernels and Libraries in the *EMR Management Guide*.
- **Managed Scaling** – With Amazon EMR version 5.30.0 and later, you can enable EMR managed scaling to automatically increase or decrease the number of instances or units in your cluster

based on workload. Amazon EMR continuously evaluates cluster metrics to make scaling decisions that optimize your clusters for cost and speed. For more information, see Scaling Cluster Resources in the *Amazon EMR Management Guide*.

- **Encrypt log files stored in Amazon S3** – With Amazon EMR version 5.30.0 and later, you can encrypt log files stored in Amazon S3 with an AWS KMS customer managed key. For more information, see Encrypt log files stored in Amazon S3 in the *Amazon EMR Management Guide*.
- **Amazon Linux 2 support** – In EMR version 5.30.0 and later, EMR usesAmazon Linux 2 OS. New custom AMIs (Amazon Machine Image) must be based on theAmazon Linux 2 AMI. For more information, see Using a Custom AMI.
- **Presto Graceful Auto Scale** – EMR clusters using 5.30.0 can be set with an auto scaling timeout period that gives Presto tasks time to finish running before their node is decommissioned. For more information, see Using Presto automatic scaling with Graceful Decommission.
- **Fleet Instance creation with new allocation strategy option** – A new allocation strategy option is available in EMR version 5.12.1 and later. It offers faster cluster provisioning, more accurate spot allocation, and less spot instance interruption. Updates to non-default EMR service roles are required. See Configure Instance Fleets.
- **sudo systemctl stop and sudo systemctl start commands** – In EMR version 5.30.0 and later, which useAmazon Linux 2 OS, EMR uses `sudo systemctl stop` and `sudo systemctl start` commands to restart services. For more information, see How do I restart a service in Amazon EMR?.

Changes, enhancements, and resolved issues

- EMR version 5.30.0 doesn't install Ganglia by default. You can explicitly select Ganglia to install when you create a cluster.
- Spark performance optimizations.
- Presto performance optimizations.
- Python 3 is the default for Amazon EMR version 5.30.0 and later.
- The default managed security group for service access in private subnets has been updated with new rules. If you use a custom security group for service access, you must include the same rules as the default managed security group. For more information, see Amazon EMR-Managed Security Group for Service Access (Private Subnets). If you use a custom service role for Amazon EMR, you must grant permission to `ec2:describeSecurityGroups` so that EMR can validate if the security groups are correctly created. If you use the `EMR_DefaultRole`, this permission is already included in the default managed policy.

Known issues

- **Lower "Max open files" limit on older AL2 [fixed in newer releases].** Amazon EMR releases: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0, and emr-6.2.0 are based on older versions ofAmazon Linux 2 (AL2), which have a lower ulimit setting for "Max open files" when Amazon EMR clusters are created with the default AMI. Amazon EMR releases 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 and later include a permanent fix with a higher "Max open files" setting. Releases with the lower open file limit causes a "Too many open files" error when submitting Spark job. In the impacted releases, the Amazon EMR default AMI has a default ulimit setting of 4096 for "Max open files," which is lower than the 65536 file limit in the latestAmazon Linux 2 AMI. The lower ulimit setting for "Max open files" causes Spark job failure when the Spark driver and executor try to open more than 4096 files. To fix the issue, Amazon EMR has a bootstrap action (BA) script that adjusts the ulimit setting at cluster creation.

  If you are using an older Amazon EMR version that doesn't have the permanent fix for this issue, the following workaround lets you to explicitly set the instance-controller ulimit to a maximum of 65536 files.

  Explicitly set a ulimit from the command line

  1. Edit `/etc/systemd/system/instance-controller.service` to add the following parameters to Service section.

     `LimitNOFILE=65536`

     `LimitNPROC=65536`
  2. Restart InstanceController

     `$ sudo systemctl daemon-reload`

     `$ sudo systemctl restart instance-controller`

  Set a ulimit using bootstrap action (BA)

  You can also use a bootstrap action (BA) script to configure the instance-controller ulimit to 65536 files at cluster creation.

```
#!/bin/bash
for user in hadoop spark hive; do
```

```
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- **Managed scaling**

  Managed scaling operations on 5.30.0 and 5.30.1 clusters without Presto installed may cause application failures or cause a uniform instance group or instance fleet to stay in the ARRESTED state, particularly when a scale down operation is followed quickly by a scale up operation.

  As a workaround, choose Presto as an application to install when you create a cluster with Amazon EMR releases 5.30.0 and 5.30.1, even if your job does not require Presto.

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as hadoop user to the lead primary node of the EMR cluster with multiple primary nodes.
  - Run the following command to renew Kerberos ticket for hadoop user.

  ```
  kinit -kt <keytab_file> <principal>
  ```

Typically, the keytab file is located at `/etc/hadoop.keytab` and the principal is in the form of `hadoop/<hostname>@<REALM>`.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

- The default database engine for Hue 4.6.0 is SQLite, which causes issues when you try to use Hue with an external database. To fix this, set `engine` in your `hue-ini` configuration classification to `mysql`. This issue has been fixed in Amazon EMR version 5.30.1.
- When you use Spark with Hive partition location formatting to read data in Amazon S3, and you run Spark on Amazon EMR releases 5.30.0 to 5.36.0, and 6.2.0 to 6.9.0, you might encounter an issue that prevents your cluster from reading data correctly. This can happen if your partitions have all of the following characteristics:
  - Two or more partitions are scanned from the same table.
  - At least one partition directory path is a prefix of at least one other partition directory path, for example, `s3://bucket/table/p=a` is a prefix of `s3://bucket/table/p=a b`.
  - The first character that follows the prefix in the other partition directory has a UTF-8 value that's less than than the / character (U+002F). For example, the space character (U+0020) that occurs between a and b in `s3://bucket/table/p=a b` falls into this category. Note that there are 14 other non-control characters: `!"#$%&‘()*+,-`. For more information, see [UTF-8 encoding table and Unicode characters](#).

  As a workaround to this issue, set the `spark.sql.sources.fastS3PartitionDiscovery.enabled` configuration to `false` in the `spark-defaults` classification.

Release 5.29.0

The following release notes include information for Amazon EMR release 5.29.0. Changes are relative to 5.28.1.

Initial release date: Jan 17, 2020

Upgrades

- Upgraded AWS SDK for Java to version 1.11.682
- Upgraded Hive to version 2.3.6
- Upgraded Flink to version 1.9.1
- Upgraded EmrFS to version 2.38.0
- Upgraded EMR DynamoDB Connector to version 4.13.0

Changes, enhancements, and resolved issues

- Spark
  - Spark performance optimizations.
- EMRFS
  - Management Guide updates to emrfs-site.xml default settings for consistent view.

Known issues

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as `hadoop` user to the lead primary node of the EMR cluster with multiple primary nodes.
  - Run the following command to renew Kerberos ticket for `hadoop` user.

    ```
    kinit -kt <keytab_file> <principal>
    ```

    Typically, the keytab file is located at `/etc/hadoop.keytab` and the principal is in the form of `hadoop/<hostname>@<REALM>`.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

Release 5.28.1

The following release notes include information for Amazon EMR release 5.28.1. Changes are relative to 5.28.0.

Initial release date: Jan 10, 2020

Changes, enhancements, and resolved issues

- Spark
  - Fixed Spark compatibility issues.
- CloudWatch Metrics
  - Fixed Amazon CloudWatch Metrics publishing on an EMR cluster with multiple primary nodes.
- Disabled log message
  - Disabled false log message, "...using old version (<4.5.8) of Apache http client."

Known issues

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as `hadoop` user to the lead primary node of the EMR cluster with multiple primary nodes.

- Run the following command to renew Kerberos ticket for hadoop user.

```
kinit -kt <keytab_file> <principal>
```

Typically, the keytab file is located at /etc/hadoop.keytab and the principal is in the form of hadoop/<hostname>@<REALM>.

Note

This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

Release 5.28.0

The following release notes include information for Amazon EMR release 5.28.0. Changes are relative to 5.27.0.

Initial release date: Nov 12, 2019

Upgrades

- Upgraded Flink to version 1.9.0
- Upgraded Hive to version 2.3.6
- Upgraded MXNet to version 1.5.1
- Upgraded Phoenix to version 4.14.3
- Upgraded Presto to version 0.227
- Upgraded Zeppelin to version 0.8.2

New features

- Apache Hudi is now available for Amazon EMR to install when you create a cluster. For more information, see Hudi.
- (Nov 25, 2019) You can now choose to run multiple steps in parallel to improve cluster utilization and save cost. You can also cancel both pending and running steps. For more information, see Work with Steps Using the AWS CLI and Console.

- (Dec 3, 2019) You can now create and run EMR clusters on AWS Outposts. AWS Outposts enables native AWS services, infrastructure, and operating models in on-premises facilities. In AWS Outposts environments, you can use the same AWS APIs, tools, and infrastructure that you use in the AWS cloud. For more information, see EMR clusters on AWS Outposts.
- (Mar 11, 2020) Beginning with Amazon EMR version 5.28.0, you can create and run Amazon EMR clusters on an AWS Local Zones subnet as a logical extension of an AWS Region that supports Local Zones. A Local Zone enables Amazon EMR features and a subset of AWS services, like compute and storage services, to be located closer to users, providing very low latency access to applications running locally. For a list of available Local Zones, see AWS Local Zones. For information about accessing available AWS Local Zones, see Regions, Availability Zones, and Local Zones.

  Local Zones do not currently support Amazon EMR Notebooks and do not support connections directly to Amazon EMR using interface VPC endpoint (AWS PrivateLink).

Changes, enhancements, and resolved issues

- Expanded Application Support for High Availability Clusters
  - For more information, see Supported applications in an EMR cluster with Multiple Primary Nodes in the *Amazon EMR Management Guide*.
- Spark
  - Performance optimizations
- Hive
  - Performance optimizations
- Presto
  - Performance optimizations

Known issues

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem

impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

Workaround:

- SSH as hadoop user to the lead primary node of the EMR cluster with multiple primary nodes.
- Run the following command to renew Kerberos ticket for hadoop user.

```
kinit -kt <keytab_file> <principal>
```

  Typically, the keytab file is located at /etc/hadoop.keytab and the principal is in the form of hadoop/<hostname>@<REALM>.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

Release 5.27.0

The following release notes include information for Amazon EMR release 5.27.0. Changes are relative to 5.26.0.

Initial release date: Sep 23, 2019

Upgrades

- AWS SDK for Java 1.11.615
- Flink 1.8.1
- JupyterHub 1.0.0
- Spark 2.4.4
- Tensorflow 1.14.0
- Connectors and drivers:
  - DynamoDB Connector 4.12.0

New features

- (Oct 24, 2019) The following New features in EMR notebooks are available with all Amazon EMR releases.
  - You can now associate Git repositories with EMR notebooks to store your notebooks in a version controlled environment. You can share code with peers and reuse existing Jupyter notebooks through remote Git repositories. For more information, see Associate Git Repositories with Amazon EMR Notebooks in the *Amazon EMR Management Guide*.
  - The nbdime utility is now available in EMR notebooks to simplify comparing and merging notebooks.
  - EMR notebooks now support JupyterLab. JupyterLab is a web-based interactive development environment fully compatible with Jupyter notebooks. You can now choose to open your notebook in either JupyterLab or Jupyter notebook editor.
- (Oct 30, 2019) With Amazon EMR versions 5.25.0 and later, you can connect to Spark history server UI from the cluster **Summary** page or the **Application history** tab in the console. Instead of setting up a web proxy through an SSH connection, you can quickly access the Spark history server UI to view application metrics and access relevant log files for active and terminated clusters. For more information, see Off-cluster access to persistent application user interfaces in the *Amazon EMR Management Guide*.

Changes, enhancements, and resolved issues

- Amazon EMR cluster with multiple primary nodes
  - You can install and run Flink on an Amazon EMR cluster with multiple primary nodes. For more information, see Supported applications and features.
  - You can configure HDFS transparent encryption on an Amazon EMR cluster with multiple primary nodes. For more information, see HDFS Transparent Encryption on EMR clusters with Multiple Primary Nodes.
  - You can now modify the configuration of applications running on an Amazon EMR cluster with multiple primary nodes. For more information, see Supplying a Configuration for an Instance Group in a Running Cluster.
- Amazon EMR-DynamoDB Connector
  - Amazon EMR-DynamoDB Connector now supports the following DynamoDB data types: boolean, list, map, item, null. For more information, see Set Up a Hive Table to Run Hive Commands.

Known issues

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as `hadoop` user to the lead primary node of the EMR cluster with multiple primary nodes.
  - Run the following command to renew Kerberos ticket for `hadoop` user.

    ```
    kinit -kt <keytab_file> <principal>
    ```

    Typically, the keytab file is located at `/etc/hadoop.keytab` and the principal is in the form of `hadoop/<hostname>@<REALM>`.

> **Note**
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

Release 5.26.0

The following release notes include information for Amazon EMR release 5.26.0. Changes are relative to 5.25.0.

Initial release date: Aug 8, 2019

Last updated date: Aug 19, 2019

Upgrades

- AWS SDK for Java 1.11.595

- HBase 1.4.10
- Phoenix 4.14.2
- Connectors and drivers:
  - DynamoDB Connector 4.11.0
  - MariaDB Connector 2.4.2
  - Amazon Redshift JDBC Driver 1.2.32.1056

New features

- (Beta) With Amazon EMR 5.26.0, you can launch a cluster that integrates with Lake Formation. This integration provides fine-grained, column-level access to databases and tables in the AWS Glue Data Catalog. It also enables federated single sign-on to EMR Notebooks or Apache Zeppelin from an enterprise identity system. For more information, see Integrating Amazon EMR with AWS Lake Formation (Beta).
- (Aug 19, 2019) Amazon EMR block public access is now available with all Amazon EMR releases that support security groups. Block public access is an account-wide setting applied to each AWS Region. Block public access prevents a cluster from launching when any security group associated with the cluster has a rule that allows inbound traffic from IPv4 0.0.0.0/0 or IPv6 ::/0 (public access) on a port, unless a port is specified as an exception. Port 22 is an exception by default. For more information, see Using Amazon EMR Block Public Access in the *Amazon EMR Management Guide*.

Changes, enhancements, and resolved issues

- EMR Notebooks
  - With EMR 5.26.0 and later, EMR Notebooks supports notebook-scoped Python libraries in addition to the default Python libraries. You can install notebook-scoped libraries from within the notebook editor without having to re-create a cluster or re-attach a notebook to a cluster. Notebook-scoped libraries are created in a Python virtual environment, so they apply only to the current notebook session. This allows you to isolate notebook dependencies. For more information, see Using Notebook Scoped Libraries in the *Amazon EMR Management Guide*.
- EMRFS
  - You can enable an ETag verification feature (Beta) by setting `fs.s3.consistent.metadata.etag.verification.enabled` to `true`. With this feature, EMRFS uses Amazon S3 ETags to verify that objects being read are the latest available

version. This feature is helpful for read-after-update use cases in which files on Amazon S3 are overwritten while retaining the same name. This ETag verification capability currently does not work with S3 Select. For more information, see Configure Consistent View.

- Spark
  - The following optimizations are now enabled by default: dynamic partition pruning, DISTINCT before INTERSECT, improvements in SQL plan statistics inference for JOIN followed by DISTINCT queries, flattening scalar subqueries, optimized join reorder, and bloom filter join. For more information, see Optimizing Spark Performance.
  - Improved whole stage code generation for Sort Merge Join.
  - Improved query fragment and subquery reuse.
  - Improvements to pre-allocate executors on Spark start up.
  - Bloom filter joins are no longer applied when the smaller side of the join includes a broadcast hint.
- Tez
  - Resolved an issue with Tez. Tez UI now works on an Amazon EMR cluster with multiple primary nodes.

Known issues

- The improved whole stage code generation capabilities for Sort Merge Join can increase memory pressure when enabled. This optimization improves performance, but may result in job retries or failures if the `spark.yarn.executor.memoryOverheadFactor` is not tuned to provide enough memory. To disable this feature, set `spark.sql.sortMergeJoinExec.extendedCodegen.enabled` to false.
- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as hadoop user to the lead primary node of the EMR cluster with multiple primary nodes.

- Run the following command to renew Kerberos ticket for hadoop user.

```
kinit -kt <keytab_file> <principal>
```

Typically, the keytab file is located at /etc/hadoop.keytab and the principal is in the form of hadoop/<hostname>@<REALM>.

Note

This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

Release 5.25.0

The following release notes include information for Amazon EMR release 5.25.0. Changes are relative to 5.24.1.

Initial release date: July 17, 2019

Last updated date: Oct 30, 2019

Amazon EMR 5.25.0

Upgrades

- AWS SDK for Java 1.11.566
- Hive 2.3.5
- Presto 0.220
- Spark 2.4.3
- TensorFlow 1.13.1
- Tez 0.9.2
- Zookeeper 3.4.14

New features

- (Oct 30, 2019) Beginning with Amazon EMR version 5.25.0, you can connect to Spark history server UI from the cluster **Summary** page or the **Application history** tab in the console. Instead of setting up a web proxy through an SSH connection, you can quickly access the Spark history server UI to view application metrics and access relevant log files for active and terminated clusters. For more information, see Off-cluster access to persistent application user interfaces in the *Amazon EMR Management Guide*.

Changes, enhancements, and resolved issues

- Spark
  - Improved the performance of some joins by using Bloom filters to pre-filter inputs. The optimization is disabled by default and can be enabled by setting the Spark configuration parameter `spark.sql.bloomFilterJoin.enabled` to `true`.
  - Improved the performance of grouping by string type columns.
  - Improved the default Spark executor memory and cores configuration of R4 instance types for clusters without HBase installed.
  - Resolved a previous issue with the dynamic partition pruning feature where the pruned table has to be on the left side of the join.
  - Improved DISTINCT before INTERSECT optimization to apply to additional cases involving aliases.
  - Improved SQL plan statistics inference for JOIN followed by DISTINCT queries. This improvement is disabled by default and can be enabled by setting the Spark configuration parameter `spark.sql.statsImprovements.enabled` to `true`. This optimization is required by the Distinct before Intersect feature and will be enabled automatically when `spark.sql.optimizer.distinctBeforeIntersect.enabled` is set to `true`.
  - Optimized join order based on table size and filters. This optimization is disabled by default and can be enabled by setting the Spark configuration parameter `spark.sql.optimizer.sizeBasedJoinReorder.enabled` to `true`.

  For more information, see Optimizing Spark Performance.
- EMRFS
  - The EMRFS setting, `fs.s3.buckets.create.enabled`, is now disabled by default. With testing, we found that disabling this setting improves performance and prevents unintentional

creation of S3 buckets. If your application relies on this functionality, you can enable it by setting the property `fs.s3.buckets.create.enabled` to `true` in the `emrfs-site` configuration classification. For information, see Supplying a Configuration when Creating a Cluster.

- Local Disk Encryption and S3 Encryption Improvements in Security Configurations (August 5, 2019)
  - Separated Amazon S3 encryption settings from local disk encryption settings in security configuration setup.
  - Added an option to enable EBS encryption with release 5.24.0 and later. Selecting this option encrypts the root device volume in addition to storage volumes. Previous versions required using a custom AMI to encrypt the root device volume.
  - For more information, see Encryption Options in the *Amazon EMR Management Guide*.

Known issues

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as `hadoop` user to the lead primary node of the EMR cluster with multiple primary nodes.
  - Run the following command to renew Kerberos ticket for `hadoop` user.

    ```
    kinit -kt <keytab_file> <principal>
    ```

    Typically, the keytab file is located at `/etc/hadoop.keytab` and the principal is in the form of `hadoop/<hostname>@<REALM>`.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

Release 5.24.1

The following release notes include information for Amazon EMR release 5.24.1. Changes are relative to 5.24.0.

Initial release date: June 26, 2019

Changes, enhancements, and resolved issues

- Updated the default Amazon Linux AMI for Amazon EMR to include important Linux kernel security updates, including the TCP SACK Denial of Service Issue (AWS-2019-005).

Known issues

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as hadoop user to the lead primary node of the EMR cluster with multiple primary nodes.
  - Run the following command to renew Kerberos ticket for hadoop user.

    ```
    kinit -kt <keytab_file> <principal>
    ```

Typically, the keytab file is located at `/etc/hadoop.keytab` and the principal is in the form of `hadoop/<hostname>@<REALM>`.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

Release 5.24.0

The following release notes include information for Amazon EMR release 5.24.0. Changes are relative to 5.23.0.

Initial release date: June 11, 2019

Last updated date: August 5, 2019

Upgrades

- Flink 1.8.0
- Hue 4.4.0
- JupyterHub 0.9.6
- Livy 0.6.0
- MxNet 1.4.0
- Presto 0.219
- Spark 2.4.2
- AWS SDK for Java 1.11.546
- Connectors and drivers:
  - DynamoDB Connector 4.9.0
  - MariaDB Connector 2.4.1
  - Amazon Redshift JDBC Driver 1.2.27.1051

Changes, enhancements, and resolved issues

- Spark
  - Added optimization to dynamically prune partitions. The optimization is disabled by default. To enable it, set the Spark configuration parameter `spark.sql.dynamicPartitionPruning.enabled` to `true`.
  - Improved performance of `INTERSECT` queries. This optimization is disabled by default. To enable it, set the Spark configuration parameter `spark.sql.optimizer.distinctBeforeIntersect.enabled` to `true`.
  - Added optimization to flatten scalar subqueries with aggregates that use the same relation. The optimization is disabled by default. To enable it, set the Spark configuration parameter `spark.sql.optimizer.flattenScalarSubqueriesWithAggregates.enabled` to `true`.
  - Improved whole stage code generation.

  For more information, see Optimizing Spark Performance.
- Local Disk Encryption and S3 Encryption Improvements in Security Configurations (August 5, 2019)
  - Separated Amazon S3 encryption settings from local disk encryption settings in security configuration setup.
  - Added an option to enable EBS encryption. Selecting this option encrypts the root device volume in addition to storage volumes. Previous versions required using a custom AMI to encrypt the root device volume.
  - For more information, see Encryption Options in the *Amazon EMR Management Guide*.

Known issues

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

Workaround:

- SSH as hadoop user to the lead primary node of the EMR cluster with multiple primary nodes.
- Run the following command to renew Kerberos ticket for hadoop user.

```
kinit -kt <keytab_file> <principal>
```

  Typically, the keytab file is located at /etc/hadoop.keytab and the principal is in the form of hadoop/<hostname>@<REALM>.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

Release 5.23.0

The following release notes include information for Amazon EMR release 5.23.0. Changes are relative to 5.22.0.

Initial release date: April 01, 2019

Last updated date: April 30, 2019

Upgrades

- AWS SDK for Java 1.11.519

New features

- (April 30, 2019) With Amazon EMR 5.23.0 and later, you can launch a cluster with three primary nodes to support high availability of applications like YARN Resource Manager, HDFS NameNode, Spark, Hive, and Ganglia. The primary node is no longer a potential single point of failure with this feature. If one of the primary nodes fails, Amazon EMR automatically fails over to a standby primary node and replaces the failed primary node with a new one with the same configuration and bootstrap actions. For more information, see Plan and Configure Primary Nodes.

Known issues

- Tez UI (Fixed in Amazon EMR release 5.26.0)

  Tez UI does not work on an EMR cluster with multiple primary nodes.
- Hue (Fixed in Amazon EMR release 5.24.0)
  - Hue running on Amazon EMR does not support Solr. Beginning with Amazon EMR release 5.20.0, a misconfiguration issue causes Solr to be enabled and a harmless error message to appear similar to the following:

    ```
    Solr server could not be contacted properly:
    HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
    port=1978): Max retries exceeded with url: /solr/admin/info/
    system?user.name=hue&doAs=administrator&wt=json (Caused by
    NewConnectionError(': Failed to establish a new connection: [Errno 111]
    Connection refused',))
    ```

    To prevent the Solr error message from appearing:

    1. Connect to the primary node command line using SSH.
    2. Use a text editor to open the `hue.ini` file. For example:

       ```
       sudo vim /etc/hue/conf/hue.ini
       ```
    3. Search for the term `appblacklist` and modify the line to the following:

       ```
       appblacklist = search
       ```
    4. Save your changes and restart Hue as shown in the following example:

       ```
       sudo stop hue; sudo start hue
       ```
- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

Workaround:

- SSH as hadoop user to the lead primary node of the EMR cluster with multiple primary nodes.
- Run the following command to renew Kerberos ticket for hadoop user.

```
kinit -kt <keytab_file> <principal>
```

Typically, the keytab file is located at /etc/hadoop.keytab and the principal is in the form of hadoop/<hostname>@<REALM>.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

Release 5.22.0

The following release notes include information for Amazon EMR release 5.22.0. Changes are relative to 5.21.0.

> **Important**
>
> Beginning with Amazon EMR release 5.22.0, Amazon EMR uses AWS Signature Version 4 exclusively to authenticate requests to Amazon S3. Earlier Amazon EMR releases use AWS Signature Version 2 in some cases, unless the release notes indicate that Signature Version 4 is used exclusively. For more information, see Authenticating Requests (AWS Signature Version 4) and Authenticating Requests (AWS Signature Version 2) in the *Amazon Simple Storage Service Developer Guide*.

Initial release date: March 20, 2019

Upgrades

- Flink 1.7.1
- HBase 1.4.9

- Oozie 5.1.0
- Phoenix 4.14.1
- Zeppelin 0.8.1
- Connectors and drivers:
  - DynamoDB Connector 4.8.0
  - MariaDB Connector 2.2.6
  - Amazon Redshift JDBC Driver 1.2.20.1043

New features

- Modified the default EBS configuration for EC2 instance types with EBS-only storage. When you create a cluster using Amazon EMR release 5.22.0 and later, the default amount of EBS storage increases based on the size of the instance. In addition, we split increased storage across multiple volumes, giving increased IOPS performance. If you want to use a different EBS instance storage configuration, you can specify it when you create an EMR cluster or add nodes to an existing cluster. For more information about the amount of storage and number of volumes allocated by default for each instance type, see Default EBS Storage for Instances in the *Amazon EMR Management Guide*.

Changes, enhancements, and resolved issues

- Spark
  - Introduced a new configuration property for Spark on YARN, `spark.yarn.executor.memoryOverheadFactor`. The value of this property is a scale factor that sets the value of memory overhead to a percentage of executor memory, with a minimum of 384 MB. If memory overhead is set explicitly using `spark.yarn.executor.memoryOverhead`, this property has no effect. The default value is `0.1875`, representing 18.75%. This default for Amazon EMR leaves more space in YARN containers for executor memory overhead than the 10% default set internally by Spark. The Amazon EMR default of 18.75% empirically showed fewer memory-related failures in TPC-DS benchmarks.
  - Backported SPARK-26316 to improve performance.
- In Amazon EMR version 5.19.0, 5.20.0, and 5.21.0, YARN node labels are stored in an HDFS directory. In some situations, this leads to core node startup delays and then causes cluster time-

out and launch failure. Beginning with Amazon EMR 5.22.0, this issue is resolved. YARN node labels are stored on the local disk of each cluster node, avoiding dependencies on HDFS.

Known issues

- Hue (Fixed in Amazon EMR release 5.24.0)
  - Hue running on Amazon EMR does not support Solr. Beginning with Amazon EMR release 5.20.0, a misconfiguration issue causes Solr to be enabled and a harmless error message to appear similar to the following:

    ```
    Solr server could not be contacted properly: 
    HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal', 
    port=1978): Max retries exceeded with url: /solr/admin/info/
    system?user.name=hue&doAs=administrator&wt=json (Caused by 
    NewConnectionError(': Failed to establish a new connection: [Errno 111] 
    Connection refused',))
    ```

    To prevent the Solr error message from appearing:

    1. Connect to the primary node command line using SSH.
    2. Use a text editor to open the `hue.ini` file. For example:

       ```
       sudo vim /etc/hue/conf/hue.ini
       ```

    3. Search for the term `appblacklist` and modify the line to the following:

       ```
       appblacklist = search
       ```

    4. Save your changes and restart Hue as shown in the following example:

       ```
       sudo stop hue; sudo start hue
       ```

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

Workaround:

- SSH as hadoop user to the lead primary node of the EMR cluster with multiple primary nodes.
- Run the following command to renew Kerberos ticket for hadoop user.

```
kinit -kt <keytab_file> <principal>
```

  Typically, the keytab file is located at /etc/hadoop.keytab and the principal is in the form of hadoop/<hostname>@<REALM>.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

Release 5.21.1

The following release notes include information for Amazon EMR release 5.21.1. Changes are relative to 5.21.0.

Initial release date: July 18, 2019

Changes, enhancements, and resolved issues

- Updated the default Amazon Linux AMI for Amazon EMR to include important Linux kernel security updates, including the TCP SACK Denial of Service Issue (AWS-2019-005).

Known issues

- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem

impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

Workaround:

- SSH as hadoop user to the lead primary node of the EMR cluster with multiple primary nodes.
- Run the following command to renew Kerberos ticket for hadoop user.

```
kinit -kt <keytab_file> <principal>
```

  Typically, the keytab file is located at /etc/hadoop.keytab and the principal is in the form of hadoop/<hostname>@<REALM>.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

Release 5.21.0

The following release notes include information for Amazon EMR release 5.21.0. Changes are relative to 5.20.0.

Initial release date: February 18, 2019

Last updated date: April 3, 2019

Upgrades

- Flink 1.7.0
- Presto 0.215
- AWS SDK for Java 1.11.479

New features

- (April 3, 2019) With Amazon EMR version 5.21.0 and later, you can override cluster configurations and specify additional configuration classifications for each instance group in a running cluster. You do this by using the Amazon EMR console, the AWS Command Line Interface (AWS CLI), or the AWS SDK. For more information, see Supplying a Configuration for an Instance Group in a Running Cluster.

Changes, enhancements, and resolved issues

- Zeppelin
  - Backported ZEPPELIN-3878.

Known issues

- Hue (Fixed in Amazon EMR release 5.24.0)
  - Hue running on Amazon EMR does not support Solr. Beginning with Amazon EMR release 5.20.0, a misconfiguration issue causes Solr to be enabled and a harmless error message to appear similar to the following:

    ```
    Solr server could not be contacted properly:
    HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
    port=1978): Max retries exceeded with url: /solr/admin/info/
    system?user.name=hue&doAs=administrator&wt=json (Caused by
    NewConnectionError(': Failed to establish a new connection: [Errno 111]
    Connection refused',))
    ```

    To prevent the Solr error message from appearing:

    1. Connect to the primary node command line using SSH.
    2. Use a text editor to open the hue.ini file. For example:

       ```
       sudo vim /etc/hue/conf/hue.ini
       ```

    3. Search for the term appblacklist and modify the line to the following:

       ```
       appblacklist = search
       ```

    4. Save your changes and restart Hue as shown in the following example:

       ```
       sudo stop hue; sudo start hue
       ```

- Tez

  - This issue was fixed in Amazon EMR 5.22.0.

    When you connect to the Tez UI at http://*MasterDNS*:8080/tez-ui through an SSH connection to the cluster primary node, the error "Adapter operation failed - Timeline server (ATS) is out of reach. Either it is down, or CORS is not enabled" appears, or tasks unexpectedly show N/A.

    This is caused by the Tez UI making requests to the YARN Timeline Server using `localhost` rather than the host name of the primary node. As a workaround, a script is available to run as a bootstrap action or step. The script updates the host name in the Tez `configs.env` file. For more information and the location of the script, see the [Bootstrap Instructions].

- In Amazon EMR version 5.19.0, 5.20.0, and 5.21.0, YARN node labels are stored in an HDFS directory. In some situations, this leads to core node startup delays and then causes cluster time-out and launch failure. Beginning with Amazon EMR 5.22.0, this issue is resolved. YARN node labels are stored on the local disk of each cluster node, avoiding dependencies on HDFS.
- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

  Workaround:

  - SSH as `hadoop` user to the lead primary node of the EMR cluster with multiple primary nodes.
  - Run the following command to renew Kerberos ticket for `hadoop` user.

    ```
    kinit -kt <keytab_file> <principal>
    ```

    Typically, the keytab file is located at `/etc/hadoop.keytab` and the principal is in the form of `hadoop/<hostname>@<REALM>`.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

Release 5.20.0

The following release notes include information for Amazon EMR release 5.20.0. Changes are relative to 5.19.0.

Initial release date: December 18, 2018

Last updated date: January 22, 2019

Upgrades

- Flink 1.6.2
- HBase 1.4.8
- Hive 2.3.4
- Hue 4.3.0
- MXNet 1.3.1
- Presto 0.214
- Spark 2.4.0
- TensorFlow 1.12.0
- Tez 0.9.1
- AWS SDK for Java 1.11.461

New features

- (January 22, 2019) Kerberos in Amazon EMR has been improved to support authenticating principals from an external KDC. This centralizes principal management because multiple clusters can share a single, external KDC. In addition, the external KDC can have a cross-realm trust with an Active Directory domain. This allows all clusters to authenticate principals from

Active Directory. For more information, see Use Kerberos Authentication in the *Amazon EMR Management Guide*.

Changes, enhancements, and resolved issues

- Default Amazon Linux AMI for Amazon EMR
  - Python3 package was upgraded from python 3.4 to 3.6.
- The EMRFS S3-optimized committer
  - The EMRFS S3-optimized committer is now enabled by default, which improves write performance. For more information, see Use the EMRFS S3-optimized committer.
- Hive
  - Backported HIVE-16686.
- Glue with Spark and Hive
  - In EMR 5.20.0 or later, parallel partition pruning is enabled automatically for Spark and Hive when AWS Glue Data Catalog is used as the metastore. This change significantly reduces query planning time by executing multiple requests in parallel to retrieve partitions. The total number of segments that can be executed concurrently range between 1 and 10. The default value is 5, which is a recommended setting. You can change it by specifying the property `aws.glue.partition.num.segments` in `hive-site` configuration classification. If throttling occurs, you can turn off the feature by changing the value to 1. For more information, see AWS Glue Segment Structure.

Known issues

- Hue (Fixed in Amazon EMR release 5.24.0)
  - Hue running on Amazon EMR does not support Solr. Beginning with Amazon EMR release 5.20.0, a misconfiguration issue causes Solr to be enabled and a harmless error message to appear similar to the following:

    ```
    Solr server could not be contacted properly:
    HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
    port=1978): Max retries exceeded with url: /solr/admin/info/
    system?user.name=hue&doAs=administrator&wt=json (Caused by
    NewConnectionError(': Failed to establish a new connection: [Errno 111]
    Connection refused',))
    ```

To prevent the Solr error message from appearing:

1. Connect to the primary node command line using SSH.
2. Use a text editor to open the `hue.ini` file. For example:

   `sudo vim /etc/hue/conf/hue.ini`
3. Search for the term `appblacklist` and modify the line to the following:

```
appblacklist = search
```

4. Save your changes and restart Hue as shown in the following example:

```
sudo stop hue; sudo start hue
```

- Tez
  - This issue was fixed in Amazon EMR 5.22.0.

    When you connect to the Tez UI at http://*MasterDNS*:8080/tez-ui through an SSH connection to the cluster primary node, the error "Adapter operation failed - Timeline server (ATS) is out of reach. Either it is down, or CORS is not enabled" appears, or tasks unexpectedly show N/A.

    This is caused by the Tez UI making requests to the YARN Timeline Server using `localhost` rather than the host name of the primary node. As a workaround, a script is available to run as a bootstrap action or step. The script updates the host name in the Tez `configs.env` file. For more information and the location of the script, see the Bootstrap Instructions.
- In Amazon EMR version 5.19.0, 5.20.0, and 5.21.0, YARN node labels are stored in an HDFS directory. In some situations, this leads to core node startup delays and then causes cluster time-out and launch failure. Beginning with Amazon EMR 5.22.0, this issue is resolved. YARN node labels are stored on the local disk of each cluster node, avoiding dependencies on HDFS.
- Known issue in clusters with multiple primary nodes and Kerberos authentication

  If you run clusters with multiple primary nodes and Kerberos authentication in Amazon EMR releases 5.20.0 and later, you may encounter problems with cluster operations such as scale down or step submission, after the cluster has been running for some time. The time period depends on the Kerberos ticket validity period that you defined. The scale-down problem impacts both automatic scale-down and explicit scale down requests that you submitted. Additional cluster operations can also be impacted.

Workaround:

- SSH as `hadoop` user to the lead primary node of the EMR cluster with multiple primary nodes.
- Run the following command to renew Kerberos ticket for `hadoop` user.

```
kinit -kt <keytab_file> <principal>
```

Typically, the keytab file is located at `/etc/hadoop.keytab` and the principal is in the form of `hadoop/<hostname>@<REALM>`.

> **Note**
>
> This workaround will be effective for the time period the Kerberos ticket is valid. This duration is 10 hours by default, but can configured by your Kerberos settings. You must re-run the above command once the Kerberos ticket expires.

Release 5.19.0

The following release notes include information for Amazon EMR release 5.19.0. Changes are relative to 5.18.0.

Initial release date: November 7, 2018

Last updated date: November 19, 2018

Upgrades

- Hadoop 2.8.5
- Flink 1.6.1
- JupyterHub 0.9.4
- MXNet 1.3.0
- Presto 0.212
- TensorFlow 1.11.0
- Zookeeper 3.4.13
- AWS SDK for Java 1.11.433

New features

- (Nov. 19, 2018) EMR Notebooks is a managed environment based on Jupyter Notebook. It supports Spark magic kernels for PySpark, Spark SQL, Spark R, and Scala. EMR Notebooks can be used with clusters created using Amazon EMR release 5.18.0 and later. For more information, see Using EMR Notebooks in the *Amazon EMR Management Guide*.
- The EMRFS S3-optimized committer is available when writing Parquet files using Spark and EMRFS. This committer improves write performance. For more information, see Use the EMRFS S3-optimized committer.

Changes, enhancements, and resolved issues

- YARN
  - Modified the logic that limits the application master process to running on core nodes. This functionality now uses the YARN node labels feature and properties in the `yarn-site` and `capacity-scheduler` configuration classifications. For information, see https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-instances-guidelines.html#emr-plan-spot-YARN.
- Default Amazon Linux AMI for Amazon EMR
  - `ruby18`, `php56`, and `gcc48` are no longer installed by default. These can be installed if desired using `yum`.
  - The aws-sdk ruby gem is no longer installed by default. It can be installed using `gem install aws-sdk`, if desired. Specific components can also be installed. For example, `gem install aws-sdk-s3`.

Known issues

- **EMR Notebooks**—In some circumstances, with multiple notebook editors open, the notebook editor may appear unable to connect to the cluster. If this happens, clear browser cookies and then reopen notebook editors.
- **CloudWatch ContainerPending Metric and Automatic Scaling**—(Fixed in 5.20.0)Amazon EMR may emit a negative value for `ContainerPending`. If `ContainerPending` is used in an automatic scaling rule, automatic scaling does not behave as expected. Avoid using `ContainerPending` with automatic scaling.

- In Amazon EMR version 5.19.0, 5.20.0, and 5.21.0, YARN node labels are stored in an HDFS directory. In some situations, this leads to core node startup delays and then causes cluster time-out and launch failure. Beginning with Amazon EMR 5.22.0, this issue is resolved. YARN node labels are stored on the local disk of each cluster node, avoiding dependencies on HDFS.

Release 5.18.0

The following release notes include information for Amazon EMR release 5.18.0. Changes are relative to 5.17.0.

Initial release date: October 24, 2018

Upgrades

- Flink 1.6.0
- HBase 1.4.7
- Presto 0.210
- Spark 2.3.2
- Zeppelin 0.8.0

New features

- Beginning with Amazon EMR 5.18.0, you can use the Amazon EMR artifact repository to build your job code against the exact versions of libraries and dependencies that are available with specific Amazon EMR releases. For more information, see Checking dependencies using the Amazon EMR artifact repository.

Changes, enhancements, and resolved issues

- Hive
  - Added support for S3 Select. For more information, see Using S3 Select with Hive to improve performance.
- Presto
  - Added support for S3 Select Pushdown. For more information, see Using S3 Select Pushdown with Presto to improve performance.
- Spark

  - The default log4j configuration for Spark has been changed to roll container logs hourly for Spark streaming jobs. This helps prevent the deletion of logs for long-running Spark streaming jobs.

Release 5.17.1

The following release notes include information for Amazon EMR release 5.17.1. Changes are relative to 5.17.0.

Initial release date: July 18, 2019

Changes, enhancements, and resolved issues

- Updated the default Amazon Linux AMI for Amazon EMR to include important Linux kernel security updates, including the TCP SACK Denial of Service Issue (AWS-2019-005).

Release 5.17.0

The following release notes include information for Amazon EMR release 5.17.0. Changes are relative to 5.16.0.

Initial release date: August 30, 2018

Upgrades

- Flink 1.5.2
- HBase 1.4.6
- Presto 0.206

New features

- Added support for Tensorflow. For more information, see TensorFlow.

Changes, enhancements, and resolved issues

- JupyterHub
  - Added support for notebook persistence in Amazon S3. For more information, see Configuring persistence for notebooks in Amazon S3.

- Spark
  - Added support for S3 Select. For more information, see Use S3 Select with Spark to improve query performance.
- Resolved the issues with the Cloudwatch metrics and the automatic scaling feature in Amazon EMR version 5.14.0, 5.15.0, or 5.16.0.

Known issues

- When you create a kerberized cluster with Livy installed, Livy fails with an error that simple authentication is not enabled. Rebooting the Livy server resolves the issue. As a workaround, add a step during cluster creation that runs `sudo restart livy-server` on the primary node.
- If you use a custom Amazon Linux AMI based on an Amazon Linux AMI with a creation date of 2018-08-11, the Oozie server fails to start. If you use Oozie, create a custom AMI based on an Amazon Linux AMI ID with a different creation date. You can use the following AWS CLI command to return a list of Image IDs for all HVM Amazon Linux AMIs with a 2018.03 version, along with the release date, so that you can choose an appropriate Amazon Linux AMI as your base. Replace MyRegion with your Region identifier, such as us-west-2.

```
aws ec2 --region MyRegion describe-images --owner amazon --query 'Images[?
Name!=`null`]|[?starts_with(Name, `amzn-ami-hvm-2018.03`) == `true`].
[CreationDate,ImageId,Name]' --output text | sort -rk1
```

Release 5.16.0

The following release notes include information for Amazon EMR release 5.16.0. Changes are relative to 5.15.0.

Initial release date: July 19, 2018

Upgrades

- Hadoop 2.8.4
- Flink 1.5.0
- Livy 0.5.0
- MXNet 1.2.0
- Phoenix 4.14.0

- Presto 0.203
- Spark 2.3.1
- AWS SDK for Java 1.11.336
- CUDA 9.2
- Redshift JDBC Driver 1.2.15.1025

Changes, enhancements, and resolved issues

- HBase
  - Backported HBASE-20723
- Presto
  - Configuration changes to support LDAP authentication. For more information, see Using LDAP authentication for Presto on Amazon EMR.
- Spark
  - Apache Spark version 2.3.1, available beginning with Amazon EMR release 5.16.0, addresses CVE-2018-8024 and CVE-2018-1334. We recommend that you migrate earlier versions of Spark to Spark version 2.3.1 or later.

Known issues

- This release version does not support the c1.medium or m1.small instance types. Clusters using either of these instance types fail to start. As a workaround, specify a different instance type or use a different release version.
- When you create a kerberized cluster with Livy installed, Livy fails with an error that simple authentication is not enabled. Rebooting the Livy server resolves the issue. As a workaround, add a step during cluster creation that runs `sudo restart livy-server` on the primary node.
- After the primary node reboots or the instance controller restarts, the CloudWatch metrics will not be collected and the automatic scaling feature will not be available in Amazon EMR version 5.14.0, 5.15.0, or 5.16.0. This issue is fixed in Amazon EMR 5.17.0.

Release 5.15.0

The following release notes include information for Amazon EMR release 5.15.0. Changes are relative to 5.14.0.

Initial release date: June 21, 2018

Upgrades

- Upgraded HBase to 1.4.4
- Upgraded Hive to 2.3.3
- Upgraded Hue to 4.2.0
- Upgraded Oozie to 5.0.0
- Upgraded Zookeeper to 3.4.12
- Upgraded AWS SDK to 1.11.333

Changes, enhancements, and resolved issues

- Hive
  - Backported HIVE-18069
- Hue
  - Updated Hue to correctly authenticate with Livy when Kerberos is enabled. Livy is now supported when using Kerberos with Amazon EMR.
- JupyterHub
  - Updated JupyterHub so that Amazon EMR installs LDAP client libraries by default.
  - Fixed an error in the script that generates self-signed certificates.

Known issues

- This release version does not support the c1.medium or m1.small instance types. Clusters using either of these instance types fail to start. As a workaround, specify a different instance type or use a different release version.
- After the primary node reboots or the instance controller restarts, the CloudWatch metrics will not be collected and the automatic scaling feature will not be available in Amazon EMR version 5.14.0, 5.15.0, or 5.16.0. This issue is fixed in Amazon EMR 5.17.0.

Release 5.14.1

The following release notes include information for Amazon EMR release 5.14.1. Changes are relative to 5.14.0.

Initial release date: October 17, 2018

Updated the default AMI for Amazon EMR to address potential security vulnerabilities.

Release 5.14.0

The following release notes include information for Amazon EMR release 5.14.0. Changes are relative to 5.13.0.

Initial release date: June 4, 2018

Upgrades

- Upgraded Apache Flink to 1.4.2
- Upgraded Apache MXnet to 1.1.0
- Upgraded Apache Sqoop to 1.4.7

New features

- Added JupyterHub support. For more information, see JupyterHub.

Changes, enhancements, and resolved issues

- EMRFS
  - The userAgent string in requests to Amazon S3 has been updated to contain the user and group information of the invoking principal. This can be used with AWS CloudTrail logs for more comprehensive request tracking.
- HBase
  - Included HBASE-20447, which addresses an issue that could cause cache issues, especially with split Regions.
- MXnet
  - Added OpenCV libraries.
- Spark
  - When Spark writes Parquet files to an Amazon S3 location using EMRFS, the FileOutputCommitter algorithm has been updated to use version 2 instead of version 1. This reduces the number of renames, which improves application performance. This change does not affect:

  - Applications other than Spark.
  - Applications that write to other file systems, such as HDFS (which still use version 1 of FileOutputCommitter).
  - Applications that use other output formats, such as text or csv, that already use EMRFS direct write.

Known issues

- JupyterHub
  - Using configuration classifications to set up JupyterHub and individual Jupyter notebooks when you create a cluster is not supported. Edit the jupyterhub_config.py file and jupyter_notebook_config.py files for each user manually. For more information, see Configuring JupyterHub.
  - JupyterHub fails to start on clusters within a private subnet, failing with the message `Error: ENOENT: no such file or directory, open '/etc/jupyter/conf/server.crt'` . This is caused by an error in the script that generates self-signed certificates. Use the following workaround to generate self-signed certificates. All commands are executed while connected to the primary node.
    1. Copy the certificate generation script from the container to the primary node:

       ```
       sudo docker cp jupyterhub:/tmp/gen_self_signed_cert.sh ./
       ```

    2. Use a text editor to change line 23 to change public hostname to local hostname as shown below:

       ```
       local hostname=$(curl -s $EC2_METADATA_SERVICE_URI/local-hostname)
       ```

    3. Run the script to generate self-signed certificates:

       ```
       sudo bash ./gen_self_signed_cert.sh
       ```

    4. Move the certificate files that the script generates to the `/etc/jupyter/conf/` directory:

       ```
       sudo mv /tmp/server.crt /tmp/server.key /etc/jupyter/conf/
       ```

    You can `tail` the `jupyter.log` file to verify that JupyterHub restarted and is returning a 200 response code. For example:

```
tail -f /var/log/jupyter/jupyter.log
```

This should return a response similar to the following:

```
# [I 2018-06-14 18:56:51.356 JupyterHub app:1581] JupyterHub is now running at
 https://:9443/
# 19:01:51.359 - info: [ConfigProxy] 200 GET /api/routes
```

- After the primary node reboots or the instance controller restarts, the CloudWatch metrics will not be collected and the automatic scaling feature will not be available in Amazon EMR version 5.14.0, 5.15.0, or 5.16.0. This issue is fixed in Amazon EMR 5.17.0.

Release 5.13.0

The following release notes include information for the Amazon EMR release 5.13.0. Changes are relative to 5.12.0.

Upgrades

- Upgraded Spark to 2.3.0
- Upgraded HBase to 1.4.2
- Upgraded Presto to 0.194
- Upgraded AWS SDK for Java to 1.11.297

Changes, enhancements, and resolved issues

- Hive
  - Backported HIVE-15436. Enhanced Hive APIs to return only views.

Known issues

- MXNet does not currently have OpenCV libraries.

Release 5.12.2

The following release notes include information for Amazon EMR release 5.12.2. Changes are relative to 5.12.1.

Initial release date: August 29, 2018

Changes, enhancements, and resolved issues

- This release addresses a potential security vulnerability.

Release 5.12.1

The following release notes include information for Amazon EMR release 5.12.1. Changes are relative to 5.12.0.

Initial release date: March 29, 2018

Changes, enhancements, and resolved issues

- Updated the Amazon Linux kernel of the defaultAmazon Linux AMI for Amazon EMR to address potential vulnerabilities.

Release 5.12.0

The following release notes include information for the Amazon EMR release 5.12.0. Changes are relative to 5.11.1.

Upgrades

- AWS SDK for Java 1.11.238 ⇒ 1.11.267. For more information, see the AWS SDK for Java Change Log on GitHub.
- Hadoop 2.7.3 ⇒ 2.8.3. For more information, see Apache Hadoop Releases.
- Flink 1.3.2 ⇒ 1.4.0. For more information, see the Apache Flink 1.4.0 Release Announcement.
- HBase 1.3.1 ⇒ 1.4.0. For more information, see the HBase Release Announcement.
- Hue 4.0.1 ⇒ 4.1.0. For more information, see the Release Notes.
- MxNet 0.12.0 ⇒ 1.0.0. For more information, see the MXNet Change Log on GitHub.

- Presto 0.187 ⇒ 0.188. For more information, see the Release Notes.

Changes, enhancements, and resolved issues

- **Hadoop**
  - The `yarn.resourcemanager.decommissioning.timeout` property has changed to `yarn.resourcemanager.nodemanager-graceful-decommission-timeout-secs`. You can use this property to customize cluster scale-down. For more information, see Cluster Scale-Down in the *Amazon EMR Management Guide*.
  - The Hadoop CLI added the `-d` option to the `cp` (copy) command, which specifies direct copy. You can use this to avoid creating an intermediary `.COPYING` file, which makes copying data between Amazon S3 faster. For more information, see HADOOP-12384.
- **Pig**
  - Added the `pig-env` configuration classification, which simplifies the configuration of Pig environment properties. For more information, see Configure applications.
- **Presto**
  - Added the `presto-connector-redshift` configuration classification, which you can use to configure values in the Presto `redshift.properties` configuration file. For more information, see Redshift Connector in Presto documentation, and Configure applications.
  - Presto support for EMRFS has been added and is the default configuration. Earlier Amazon EMR releases used PrestoS3FileSystem, which was the only option. For more information, see EMRFS and PrestoS3FileSystem configuration.

    > **Note**
    >
    > If you query underlying data in Amazon S3 with Amazon EMR version 5.12.0, Presto errors can occur. This is because Presto fails to pick up configuration classification values from `emrfs-site.xml`. As a workaround, create an `emrfs` subdirectory under `usr/lib/presto/plugin/hive-hadoop2/` and create a symlink in `usr/lib/presto/plugin/hive-hadoop2/emrfs` to the existing `/usr/share/aws/emr/emrfs/conf/emrfs-site.xml` file. Then restart the presto-server process (`sudo presto-server stop` followed by `sudo presto-server start`).
- **Spark**
  - Backported SPARK-22036: BigDecimal multiplication sometimes returns null.

Known issues

- MXNet does not include OpenCV libraries.
- SparkR is not available for clusters created using a custom AMI because R is not installed by default on cluster nodes.

Release 5.11.3

The following release notes include information for Amazon EMR release 5.11.3. Changes are relative to 5.11.2.

Initial release date: July 18, 2019

Changes, enhancements, and resolved issues

- Updated the default Amazon Linux AMI for Amazon EMR to include important Linux kernel security updates, including the TCP SACK Denial of Service Issue (AWS-2019-005).

Release 5.11.2

The following release notes include information for Amazon EMR release 5.11.2. Changes are relative to 5.11.1.

Initial release date: August 29, 2018

Changes, enhancements, and resolved issues

- This release addresses a potential security vulnerability.

Release 5.11.1

The following release notes include information for the Amazon EMR version 5.11.1 release. Changes are relative to the Amazon EMR 5.11.0 release.

Initial release date: January 22, 2018

Changes, enhancements, and resolved issues

- Updated the Amazon Linux kernel of the defaultAmazon Linux AMI for Amazon EMR to address vulnerabilities associated with speculative execution (CVE-2017-5715, CVE-2017-5753, and

CVE-2017-5754). For more information, see https://aws.amazon.com/security/security-bulletins/AWS-2018-013/.

Known issues

- MXNet does not include OpenCV libraries.
- Hive 2.3.2 sets `hive.compute.query.using.stats=true` by default. This causes queries to get data from existing statistics rather than directly from data, which could be confusing. For example, if you have a table with `hive.compute.query.using.stats=true` and upload new files to the table `LOCATION`, running a `SELECT COUNT(*)` query on the table returns the count from the statistics, rather than picking up the added rows.

  As a workaround, use the `ANALYZE TABLE` command to gather new statistics, or set `hive.compute.query.using.stats=false`. For more information, see Statistics in Hive in the Apache Hive documentation.

Release 5.11.0

The following release notes include information for the Amazon EMR version 5.11.0 release. Changes are relative to the Amazon EMR 5.10.0 release.

Upgrades

The following applications and components have been upgraded in this release to include the following versions.

- Hive 2.3.2
- Spark 2.2.1
- SDK for Java 1.11.238

New features

- **Spark**
  - Added `spark.decommissioning.timeout.threshold` setting, which improves Spark decommissioning behavior when using Spot instances. For more information, see Configuring node decommissioning behavior.

- Added the aws-sagemaker-spark-sdk component to Spark, which installs Amazon SageMaker Spark and associated dependencies for Spark integration with Amazon SageMaker. You can use Amazon SageMaker Spark to construct Spark machine learning (ML) pipelines using Amazon SageMaker stages. For more information, see the SageMaker Spark readme on GitHub and Using Apache Spark with Amazon SageMaker in the *Amazon SageMaker Developer Guide*.

Known issues

- MXNet does not include OpenCV libraries.
- Hive 2.3.2 sets hive.compute.query.using.stats=true by default. This causes queries to get data from existing statistics rather than directly from data, which could be confusing. For example, if you have a table with hive.compute.query.using.stats=true and upload new files to the table LOCATION, running a SELECT COUNT(*) query on the table returns the count from the statistics, rather than picking up the added rows.

  As a workaround, use the ANALYZE TABLE command to gather new statistics, or set hive.compute.query.using.stats=false. For more information, see Statistics in Hive in the Apache Hive documentation.

Release 5.10.0

The following release notes include information for the Amazon EMR version 5.10.0 release. Changes are relative to the Amazon EMR 5.9.0 release.

Upgrades

The following applications and components have been upgraded in this release to include the following versions.

- AWS SDK for Java 1.11.221
- Hive 2.3.1
- Presto 0.187

New features

- Added support for Kerberos authentication. For more information, see Use Kerberos authentication in the *Amazon EMR Management Guide*
- Added support for IAM roles for EMRFS requests to Amazon S3. For more information, see Configure IAM roles for EMRFS requests to Amazon S3 in the *Amazon EMR Management Guide*.
- Added support for GPU-based P2 and P3 instance types. For more information, see Amazon EC2 P2 instances and Amazon EC2 P3 instances. NVIDIA driver 384.81 and CUDA driver 9.0.176 are installed on these instance types by default.
- Added support for Apache MXNet.

Changes, enhancements, and resolved issues

- Presto
  - Added support for using the AWS Glue Data Catalog as the default Hive metastore. For more information, see Using Presto with the AWS Glue Data Catalog.
  - Added support for geospatial functions.
  - Added spill to disk support for joins.
  - Added support for the Redshift connector.
- Spark
  - Backported SPARK-20640, which makes the rpc timeout and the retries for shuffle registration values configurable using `spark.shuffle.registration.timeout` and `spark.shuffle.registration.maxAttempts` properties.
  - Backported SPARK-21549, which corrects an error that occurs when writing custom OutputFormat to non-HDFS locations.
- Backported Hadoop-13270
- The Numpy, Scipy, and Matplotlib libraries have been removed from the base Amazon EMR AMI. If these libraries are required for your application, they are available in the application repository, so you can use a bootstrap action to install them on all nodes using `yum install`.
- The Amazon EMR base AMI no longer has application RPM packages included, so the RPM packages are no longer present on cluster nodes. Custom AMIs and the Amazon EMR base AMI now reference the RPM package repository in Amazon S3.

- Because of the introduction of per-second billing in Amazon EC2, the default **Scale down behavior** is now **Terminate at task completion** rather than **Terminate at instance hour**. For more information, see Configure cluster scale-down.

Known issues

- MXNet does not include OpenCV libraries.
- Hive 2.3.1 sets `hive.compute.query.using.stats=true` by default. This causes queries to get data from existing statistics rather than directly from data, which could be confusing. For example, if you have a table with `hive.compute.query.using.stats=true` and upload new files to the table `LOCATION`, running a `SELECT COUNT(*)` query on the table returns the count from the statistics, rather than picking up the added rows.

  As a workaround, use the `ANALYZE TABLE` command to gather new statistics, or set `hive.compute.query.using.stats=false`. For more information, see Statistics in Hive in the Apache Hive documentation.

Release 5.9.0

The following release notes include information for the Amazon EMR version 5.9.0 release. Changes are relative to the Amazon EMR 5.8.0 release.

Release date: October 5, 2017

Latest feature update: October 12, 2017

Upgrades

The following applications and components have been upgraded in this release to include the following versions.

- AWS SDK for Java version 1.11.183
- Flink 1.3.2
- Hue 4.0.1
- Pig 0.17.0
- Presto 0.184

New features

- Added Livy support (version 0.4.0-incubating). For more information, see Apache Livy.
- Added support for Hue Notebook for Spark.
- Added support for i3-series Amazon EC2 instances (October 12, 2017).

Changes, enhancements, and resolved issues

- Spark
  - Added a new set of features that help ensure Spark handles node termination because of a manual resize or an automatic scaling policy request more gracefully. For more information, see Configuring node decommissioning behavior.
  - SSL is used instead of 3DES for in-transit encryption for the block transfer service, which enhances performance when using Amazon EC2 instance types with AES-NI.
  - Backported SPARK-21494.
- Zeppelin
  - Backported ZEPPELIN-2377.
- HBase
  - Added patch HBASE-18533, which allows additional values for HBase BucketCache configuration using the `hbase-site` configuration classification.
- Hue
  - Added AWS Glue Data Catalog support for the Hive query editor in Hue.
  - By default, superusers in Hue can access all files that Amazon EMR IAM roles are allowed to access. Newly created users do not automatically have permissions to access the Amazon S3 filebrowser and must have the `filebrowser.s3_access` permissions enabled for their group.
- Resolved an issue that caused underlying JSON data created using AWS Glue Data Catalog to be inaccessible.

Known issues

- Cluster launch fails when all applications are installed and the default Amazon EBS root volume size is not changed. As a workaround, use the `aws emr create-cluster` command from the AWS CLI and specify a larger `--ebs-root-volume-size` parameter.

- Hive 2.3.0 sets `hive.compute.query.using.stats=true` by default. This causes queries to get data from existing statistics rather than directly from data, which could be confusing. For example, if you have a table with `hive.compute.query.using.stats=true` and upload new files to the table `LOCATION`, running a `SELECT COUNT(*)` query on the table returns the count from the statistics, rather than picking up the added rows.

  As a workaround, use the `ANALYZE TABLE` command to gather new statistics, or set `hive.compute.query.using.stats=false`. For more information, see Statistics in Hive in the Apache Hive documentation.

Release 5.8.2

The following release notes include information for Amazon EMR release 5.8.2. Changes are relative to 5.8.1.

Initial release date: March 29, 2018

Changes, enhancements, and resolved issues

- Updated the Amazon Linux kernel of the defaultAmazon Linux AMI for Amazon EMR to address potential vulnerabilities.

Release 5.8.1

The following release notes include information for the Amazon EMR version 5.8.1 release. Changes are relative to the Amazon EMR 5.8.0 release.

Initial release date: January 22, 2018

Changes, enhancements, and resolved issues

- Updated the Amazon Linux kernel of the defaultAmazon Linux AMI for Amazon EMR to address vulnerabilities associated with speculative execution (CVE-2017-5715, CVE-2017-5753, and CVE-2017-5754). For more information, see https://aws.amazon.com/security/security-bulletins/AWS-2018-013/.

Release 5.8.0

The following release notes include information for the Amazon EMR version 5.8.0 release. Changes are relative to the Amazon EMR 5.7.0 release.

Initial release date: August 10, 2017

Latest feature update: September 25, 2017

Upgrades

The following applications and components have been upgraded in this release to include the following versions:

- AWS SDK 1.11.160
- Flink 1.3.1
- Hive 2.3.0. For more information, see Release notes on the Apache Hive site.
- Spark 2.2.0. For more information, see Release notes on the Apache Spark site.

New features

- Added support for viewing application history (September 25, 2017). For more information, see Viewing application history in the *Amazon EMR Management Guide*.

Changes, enhancements, and resolved issues

- **Integration with AWS Glue Data Catalog**
  - Added ability for Hive and Spark SQL to use AWS Glue Data Catalog as the Hive metadata store. For more information, see Using the AWS Glue Data Catalog as the metastore for Hive and Use the AWS Glue Data Catalog as the metastore for Spark SQL.
- Added **Application history** to cluster details, which allows you to view historical data for YARN applications and additional details for Spark applications. For more information, see View application history in the *Amazon EMR Management Guide*.
- **Oozie**
  - Backported OOZIE-2748.
- **Hue**

  - Backported HUE-5859
- **HBase**
  - Added patch to expose the HBase master server start time through Java Management Extensions (JMX) using `getMasterInitializedTime`.
  - Added patch that improves cluster start time.

Known issues

- Cluster launch fails when all applications are installed and the default Amazon EBS root volume size is not changed. As a workaround, use the `aws emr create-cluster` command from the AWS CLI and specify a larger `--ebs-root-volume-size` parameter.
- Hive 2.3.0 sets `hive.compute.query.using.stats=true` by default. This causes queries to get data from existing statistics rather than directly from data, which could be confusing. For example, if you have a table with `hive.compute.query.using.stats=true` and upload new files to the table `LOCATION`, running a `SELECT COUNT(*)` query on the table returns the count from the statistics, rather than picking up the added rows.

  As a workaround, use the `ANALYZE TABLE` command to gather new statistics, or set `hive.compute.query.using.stats=false`. For more information, see Statistics in Hive in the Apache Hive documentation.
- **Spark**—When using Spark, there is a file handler leak issue with the apppusher daemon, which can appear for a long-running Spark job after several hours or days. To fix the issue, connect to the master node and type `sudo /etc/init.d/apppusher stop`. This stops that apppusher daemon, which Amazon EMR will restart automatically.
- **Application history**
  - Historical data for dead Spark executors is not available.
  - Application history is not available for clusters that use a security configuration to enable in-flight encryption.

Release 5.7.0

The following release notes include information for the Amazon EMR 5.7.0 release. Changes are relative to the Amazon EMR 5.6.0 release.

Release date: July 13, 2017

Upgrades

- Flink 1.3.0
- Phoenix 4.11.0
- Zeppelin 0.7.2

New features

- Added the ability to specify a custom Amazon Linux AMI when you create a cluster. For more information, see Using a custom AMI.

Changes, enhancements, and resolved issues

- **HBase**
  - Added capability to configure HBase read-replica clusters. See Using a read-replica cluster.
  - Multiple bug fixes and enhancements
- **Presto** - added ability to configure `node.properties`.
- **YARN** - added ability to configure `container-log4j.properties`
- **Sqoop** - backported SQOOP-2880, which introduces an argument that allows you to set the Sqoop temporary directory.

Release 5.6.0

The following release notes include information for the Amazon EMR 5.6.0 release. Changes are relative to the Amazon EMR 5.5.0 release.

Release date: June 5, 2017

Upgrades

- Flink 1.2.1
- HBase 1.3.1
- Mahout 0.13.0. This is the first version of Mahout to support Spark 2.x in Amazon EMR version 5.0 and later.
- Spark 2.1.1

Changes, enhancements, and resolved issues

- **Presto**
  - Added the ability to enable SSL/TLS secured communication between Presto nodes by enabling in-transit encryption using a security configuration. For more information, see In-transit data encryption.
  - Backported Presto 7661, which adds the VERBOSE option to the EXPLAIN ANALYZE statement to report more detailed, low level statistics about a query plan.

Release 5.5.3

The following release notes include information for Amazon EMR release 5.5.3. Changes are relative to 5.5.2.

Initial release date: August 29, 2018

Changes, enhancements, and resolved issues

- This release addresses a potential security vulnerability.

Release 5.5.2

The following release notes include information for Amazon EMR release 5.5.2. Changes are relative to 5.5.1.

Initial release date: March 29, 2018

Changes, enhancements, and resolved issues

- Updated the Amazon Linux kernel of the defaultAmazon Linux AMI for Amazon EMR to address potential vulnerabilities.

Release 5.5.1

The following release notes include information for the Amazon EMR 5.5.1 release. Changes are relative to the Amazon EMR 5.5.0 release.

Initial release date: January 22, 2018

Changes, enhancements, and resolved issues

- Updated the Amazon Linux kernel of the defaultAmazon Linux AMI for Amazon EMR to address vulnerabilities associated with speculative execution (CVE-2017-5715, CVE-2017-5753, and CVE-2017-5754). For more information, see https://aws.amazon.com/security/security-bulletins/AWS-2018-013/.

Release 5.5.0

The following release notes include information for the Amazon EMR 5.5.0 release. Changes are relative to the Amazon EMR 5.4.0 release.

Release date: April 26, 2017

Upgrades

- Hue 3.12
- Presto 0.170
- Zeppelin 0.7.1
- ZooKeeper 3.4.10

Changes, enhancements, and resolved issues

- **Spark**
  - Backported Spark Patch (SPARK-20115) fix DAGScheduler to recompute all the lost shuffle blocks when external shuffle service is unavailable to version 2.1.0 of Spark, which is included in this release.
- **Flink**
  - Flink is now built with Scala 2.11. If you use the Scala API and libraries, we recommend that you use Scala 2.11 in your projects.
  - Addressed an issue where `HADOOP_CONF_DIR` and `YARN_CONF_DIR` defaults were not properly set, so `start-scala-shell.sh` failed to work. Also added the ability to set these values using `env.hadoop.conf.dir` and `env.yarn.conf.dir` in `/etc/flink/conf/flink-conf.yaml` or the `flink-conf` configuration classification.
  - Introduced a new EMR-specific command, `flink-scala-shell` as a wrapper for `start-scala-shell.sh`. We recommend using this command instead of `start-scala-shell`. The

new command simplifies execution. For example, `flink-scala-shell -n 2` starts a Flink Scala shell with a task parallelism of 2.

- Introduced a new EMR-specific command, `flink-yarn-session` as a wrapper for `yarn-session.sh`. We recommend using this command instead of `yarn-session`. The new command simplifies execution. For example, `flink-yarn-session -d -n 2` starts a long-running Flink session in a detached state with two task managers.
- Addressed (FLINK-6125) commons httpclient is not shaded anymore in Flink 1.2.

- **Presto**
  - Added support for LDAP authentication. Using LDAP with Presto on Amazon EMR requires that you enable HTTPS access for the Presto coordinator (`http-server.https.enabled=true` in `config.properties`). For configuration details, see LDAP authentication in Presto documentation.
  - Added support for `SHOW GRANTS`.
- **Amazon EMR Base Linux AMI**
  - Amazon EMR releases are now based on Amazon Linux 2017.03. For more information, see Amazon Linux AMI 2017.03 release notes.
  - Removed Python 2.6 from the Amazon EMR base Linux image. Python 2.7 and 3.4 are installed by default. You can install Python 2.6 manually if necessary.

Release 5.4.0

The following release notes include information for the Amazon EMR 5.4.0 release. Changes are relative to the Amazon EMR 5.3.0 release.

Release date: March 08, 2017

Upgrades

The following upgrades are available in this release:

- Upgraded to Flink 1.2.0
- Upgraded to Hbase 1.3.0
- Upgraded to Phoenix 4.9.0

Note

If you upgrade from an earlier version of Amazon EMR to Amazon EMR version 5.4.0 or later and use secondary indexing, upgrade local indexes as described in the Apache Phoenix documentation. Amazon EMR removes the required configurations from the `hbase-site` classification, but indexes need to be repopulated. Online and offline upgrade of indexes are supported. Online upgrades are the default, which means indexes are repopulated while initializing from Phoenix clients of version 4.8.0 or greater. To specify offline upgrades, set the `phoenix.client.localIndexUpgrade` configuration to false in the `phoenix-site` classification, and then SSH to the master node to run `psql [zookeeper] -1`.

- Upgraded to Presto 0.166
- Upgraded to Zeppelin 0.7.0

Changes and enhancements

The following are changes made to Amazon EMR releases for release label emr-5.4.0:

- Added support for r4 instances. See Amazon EC2 instance types.

Release 5.3.1

The following release notes include information for the Amazon EMR 5.3.1 release. Changes are relative to the Amazon EMR 5.3.0 release.

Release date: February 7, 2017

Minor changes to backport Zeppelin patches and update the default AMI for Amazon EMR.

Release 5.3.0

The following release notes include information for the Amazon EMR 5.3.0 release. Changes are relative to the Amazon EMR 5.2.1 release.

Release date: January 26, 2017

Upgrades

The following upgrades are available in this release:

- Upgraded to Hive 2.1.1
- Upgraded to Hue 3.11.0
- Upgraded to Spark 2.1.0
- Upgraded to Oozie 4.3.0
- Upgraded to Flink 1.1.4

Changes and enhancements

The following are changes made to Amazon EMR releases for release label emr-5.3.0:

- Added a patch to Hue that allows you to use the `interpreters_shown_on_wheel` setting to configure what interpreters to show first on the Notebook selection wheel, regardless of their ordering in the `hue.ini` file.
- Added the `hive-parquet-logging` configuration classification, which you can use to configure values in Hive's `parquet-logging.properties` file.

Release 5.2.2

The following release notes include information for the Amazon EMR 5.2.2 release. Changes are relative to the Amazon EMR 5.2.1 release.

Release date: May 2, 2017

Known issues resolved from the previous releases

- Backported SPARK-194459, which addresses an issue where reading from an ORC table with char/varchar columns can fail.

Release 5.2.1

The following release notes include information for the Amazon EMR 5.2.1 release. Changes are relative to the Amazon EMR 5.2.0 release.

Release date: December 29, 2016

Upgrades

The following upgrades are available in this release:

- Upgraded to Presto 0.157.1. For more information, see Presto release notes in the Presto documentation.
- Upgraded to Zookeeper 3.4.9. For more information, see ZooKeeper release notes in the Apache ZooKeeper documentation.

Changes and enhancements

The following are changes made to Amazon EMR releases for release label emr-5.2.1:

- Added support for the Amazon EC2 m4.16xlarge instance type in Amazon EMR version 4.8.3 and later, excluding 5.0.0, 5.0.3, and 5.2.0.
- Amazon EMR releases are now based on Amazon Linux 2016.09. For more information, see https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/.
- The location of Flink and YARN configuration paths are now set by default in `/etc/default/flink` that you don't need to set the environment variables `FLINK_CONF_DIR` and `HADOOP_CONF_DIR` when running the `flink` or `yarn-session.sh` driver scripts to launch Flink jobs.
- Added support for FlinkKinesisConsumer class.

Known issues resolved from the previous releases

- Fixed an issue in Hadoop where the ReplicationMonitor thread could get stuck for a long time because of a race between replication and deletion of the same file in a large cluster.
- Fixed an issue where ControlledJob#toString failed with a null pointer exception (NPE) when job status was not successfully updated.

Release 5.2.0

The following release notes include information for the Amazon EMR 5.2.0 release. Changes are relative to the Amazon EMR 5.1.0 release.

Release date: November 21, 2016

Changes and enhancements

The following changes and enhancements are available in this release:

- Added Amazon S3 storage mode for HBase.
- Enables you to specify an Amazon S3 location for the HBase rootdir. For more information, see HBase on Amazon S3.

Upgrades

The following upgrades are available in this release:

- Upgraded to Spark 2.0.2

Known issues resolved from the previous releases

- Fixed an issue with /mnt being constrained to 2 TB on EBS-only instance types.
- Fixed an issue with instance-controller and logpusher logs being output to their corresponding .out files instead of to their normal log4j-configured .log files, which rotate hourly. The .out files don't rotate, so this would eventually fill up the /emr partition. This issue only affects hardware virtual machine (HVM) instance types.

Release 5.1.0

The following release notes include information for the Amazon EMR 5.1.0 release. Changes are relative to the Amazon EMR 5.0.0 release.

Release date: November 03, 2016

Changes and enhancements

The following changes and enhancements are available in this release:

- Added support for Flink 1.1.3.
- Presto has been added as an option in the notebook section of Hue.

Upgrades

The following upgrades are available in this release:

- Upgraded to HBase 1.2.3
- Upgraded to Zeppelin 0.6.2

Known issues resolved from the previous releases

- Fixed an issue with Tez queries on Amazon S3 with ORC files did not perform as well as earlier Amazon EMR 4.x versions.

Release 5.0.3

The following release notes include information for the Amazon EMR 5.0.3 release. Changes are relative to the Amazon EMR 5.0.0 release.

Release date: October 24, 2016

Upgrades

The following upgrades are available in this release:

- Upgraded to Hadoop 2.7.3
- Upgraded to Presto 0.152.3, which includes support for the Presto web interface. You can access the Presto web interface on the Presto coordinator using port 8889. For more information about the Presto web interface, see Web interface in the Presto documentation.
- Upgraded to Spark 2.0.1
- Amazon EMR releases are now based on Amazon Linux 2016.09. For more information, see https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/.

Release 5.0.0

Release date: July 27, 2016

Upgrades

The following upgrades are available in this release:

- Upgraded to Hive 2.1
- Upgraded to Presto 0.150
- Upgraded to Spark 2.0
- Upgraded to Hue 3.10.0
- Upgraded to Pig 0.16.0
- Upgraded to Tez 0.8.4
- Upgraded to Zeppelin 0.6.1

Changes and enhancements

The following are changes made to Amazon EMR releases for release label emr-5.0.0 or greater:

- Amazon EMR supports the latest open-source versions of Hive (version 2.1) and Pig (version 0.16.0). If you have used Hive or Pig on Amazon EMR in the past, this may affect some use cases. For more information, see Hive and Pig.
- The default execution engine for Hive and Pig is now Tez. To change this, you would edit the appropriate values in the `hive-site` and `pig-properties` configuration classifications, respectively.
- An enhanced step debugging feature was added, which allows you to see the root cause of step failures if the service can determine the cause. For more information, see Enhanced step debugging in the Amazon EMR Management Guide.
- Applications that previously ended with "-Sandbox" no longer have that suffix. This may break your automation, for example, if you are using scripts to launch clusters with these applications. The following table shows application names in Amazon EMR 4.7.2 versus Amazon EMR 5.0.0.

  Application name changes

Amazon EMR 4.7.2	Amazon EMR 5.0.0
Oozie-Sandbox	Oozie
Presto-Sandbox	Presto
Sqoop-Sandbox	Sqoop
Zeppelin-Sandbox	Zeppelin

Amazon EMR 4.7.2	Amazon EMR 5.0.0
ZooKeeper-Sandbox	ZooKeeper

- Spark is now compiled for Scala 2.11.
- Java 8 is now the default JVM. All applications run using the Java 8 runtime. There are no changes to any application's byte code target. Most applications continue to target Java 7.
- Zeppelin now includes authentication features. For more information, see Zeppelin.
- Added support for security configurations, which allow you to create and apply encryption options more easily. For more information, see Data encryption.

Release 4.9.5

The following release notes include information for Amazon EMR release 4.9.5. Changes are relative to 4.9.4.

Initial release date: August 29, 2018

Changes, enhancements, and resolved issues

- HBase
  - This release addresses a potential security vulnerability.

Release 4.9.4

The following release notes include information for Amazon EMR release 4.9.4. Changes are relative to 4.9.3.

Initial release date: March 29, 2018

Changes, enhancements, and resolved issues

- Updated the Amazon Linux kernel of the defaultAmazon Linux AMI for Amazon EMR to address potential vulnerabilities.

Release 4.9.3

The following release notes include information for the Amazon EMR 4.9.3 release. Changes are relative to the Amazon EMR 4.9.2 release.

Initial release date: January 22, 2018

Changes, enhancements, and resolved issues

- Updated the Amazon Linux kernel of the defaultAmazon Linux AMI for Amazon EMR to address vulnerabilities associated with speculative execution (CVE-2017-5715, CVE-2017-5753, and CVE-2017-5754). For more information, see https://aws.amazon.com/security/security-bulletins/AWS-2018-013/.

Release 4.9.2

The following release notes include information for the Amazon EMR 4.9.2 release. Changes are relative to the Amazon EMR 4.9.1 release.

Release date: July 13, 2017

Minor changes, bug fixes, and enhancements were made in this release.

Release 4.9.1

The following release notes include information for the Amazon EMR 4.9.1 release. Changes are relative to the Amazon EMR 4.8.4 release.

Release date: April 10, 2017

Known issues resolved from the previous releases

- Backports of HIVE-9976 and HIVE-10106
- Fixed an issue in YARN where a large number of nodes (greater than 2,000) and containers (greater than 5,000) would cause an out of memory error, for example: `"Exception in thread 'main' java.lang.OutOfMemoryError"`.

Changes and enhancements

The following are changes made to Amazon EMR releases for release label emr-4.9.1:

- Amazon EMR releases are now based on Amazon Linux 2017.03. For more information, see https://aws.amazon.com/amazon-linux-ami/2017.03-release-notes/.
- Removed Python 2.6 from the Amazon EMR base Linux image. You can install Python 2.6 manually if necessary.

Release 4.8.4

The following release notes include information for the Amazon EMR 4.8.4 release. Changes are relative to the Amazon EMR 4.8.3 release.

Release date: Feb 7, 2017

Minor changes, bug fixes, and enhancements were made in this release.

Release 4.8.3

The following release notes include information for the Amazon EMR 4.8.3 release. Changes are relative to the Amazon EMR 4.8.2 release.

Release date: December 29, 2016

Upgrades

The following upgrades are available in this release:

- Upgraded to Presto 0.157.1. For more information, see Presto release notes in the Presto documentation.
- Upgraded to Spark 1.6.3. For more information, see Spark release notes in the Apache Spark documentation.
- Upgraded to ZooKeeper 3.4.9. For more information, see ZooKeeper release notes in the Apache ZooKeeper documentation.

Changes and enhancements

The following are changes made to Amazon EMR releases for release label emr-4.8.3:

- Added support for the Amazon EC2 m4.16xlarge instance type in Amazon EMR version 4.8.3 and later, excluding 5.0.0, 5.0.3, and 5.2.0.

- Amazon EMR releases are now based on Amazon Linux 2016.09. For more information, see https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/.

Known issues resolved from the previous releases

- Fixed an issue in Hadoop where the ReplicationMonitor thread could get stuck for a long time because of a race between replication and deletion of the same file in a large cluster.
- Fixed an issue where ControlledJob#toString failed with a null pointer exception (NPE) when job status was not successfully updated.

Release 4.8.2

The following release notes include information for the Amazon EMR 4.8.2 release. Changes are relative to the Amazon EMR 4.8.0 release.

Release date: October 24, 2016

Upgrades

The following upgrades are available in this release:

- Upgraded to Hadoop 2.7.3
- Upgraded to Presto 0.152.3, which includes support for the Presto web interface. You can access the Presto web interface on the Presto coordinator using port 8889. For more information about the Presto web interface, see Web interface in the Presto documentation.
- Amazon EMR releases are now based on Amazon Linux 2016.09. For more information, see https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/.

Release 4.8.0

Release date: September 7, 2016

Upgrades

The following upgrades are available in this release:

- Upgraded to HBase 1.2.2

- Upgraded to Presto-Sandbox 0.151
- Upgraded to Tez 0.8.4
- Upgraded to Zeppelin-Sandbox 0.6.1

Changes and enhancements

The following are changes made to Amazon EMR releases for release label emr-4.8.0:

- Fixed an issue in YARN where the ApplicationMaster would attempt to clean up containers that no longer exist because their instances have been terminated.
- Corrected the hive-server2 URL for Hive2 actions in the Oozie examples.
- Added support for additional Presto catalogs.
- Backported patches: HIVE-8948, HIVE-12679, HIVE-13405, PHOENIX-3116, HADOOP-12689
- Added support for security configurations, which allow you to create and apply encryption options more easily. For more information, see Data encryption.

Release 4.7.2

The following release notes include information for Amazon EMR 4.7.2.

Release date: July 15, 2016

Features

The following features are available in this release:

- Upgraded to Mahout 0.12.2
- Upgraded to Presto 0.148
- Upgraded to Spark 1.6.2
- You can now create an AWSCredentialsProvider for use with EMRFS using a URI as a parameter. For more information, see Create an AWSCredentialsProvider for EMRFS.
- EMRFS now allows users to configure a custom DynamoDB endpoint for their Consistent View metadata using the `fs.s3.consistent.dynamodb.endpoint` property in `emrfs-site.xml`.
- Added a script in `/usr/bin` called `spark-example`, which wraps `/usr/lib/spark/spark/bin/run-example` so you can run examples directly. For instance, to run the SparkPi example

that comes with the Spark distribution, you can run `spark-example SparkPi 100` from the command line or using `command-runner.jar` as a step in the API.

Known issues resolved from previous releases

- Fixed an issue where Oozie had the `spark-assembly.jar` was not in the correct location when Spark was also installed, which resulted in failure to launch Spark applications with Oozie.
- Fixed an issue with Spark Log4j-based logging in YARN containers.

Release 4.7.1

Release date: June 10, 2016

Known issues resolved from previous releases

- Fixed an issue that extended the startup time of clusters launched in a VPC with private subnets. The bug only impacted clusters launched with the Amazon EMR 4.7.0 release.
- Fixed an issue that improperly handled listing of files in Amazon EMR for clusters launched with the Amazon EMR 4.7.0 release.

Release 4.7.0

> **Important**
>
> Amazon EMR 4.7.0 is deprecated. Use Amazon EMR 4.7.1 or later instead.

Release date: June 2, 2016

Features

The following features are available in this release:

- Added Apache Phoenix 4.7.0
- Added Apache Tez 0.8.3
- Upgraded to HBase 1.2.1
- Upgraded to Mahout 0.12.0

- Upgraded to Presto 0.147
- Upgraded the AWS SDK for Java to 1.10.75
- The final flag was removed from the `mapreduce.cluster.local.dir` property in `mapred-site.xml` to allow users to run Pig in local mode.

Amazon Redshift JDBC drivers available on cluster

Amazon Redshift JDBC drivers are now included at `/usr/share/aws/redshift/jdbc`. `/usr/share/aws/redshift/jdbc/RedshiftJDBC41.jar` is the JDBC 4.1-compatible Amazon Redshift driver and `/usr/share/aws/redshift/jdbc/RedshiftJDBC4.jar` is the JDBC 4.0-compatible Amazon Redshift driver. For more information, see Configure a JDBC connection in the *Amazon Redshift Management Guide*.

Java 8

Except for Presto, OpenJDK 1.7 is the default JDK used for all applications. However, both OpenJDK 1.7 and 1.8 are installed. For information about how to set `JAVA_HOME` for applications, see Configuring applications to use Java 8.

Known issues resolved from previous releases

- Fixed a kernel issue that significantly affected performance on Throughput Optimized HDD (st1) EBS volumes for Amazon EMR in emr-4.6.0.
- Fixed an issue where a cluster would fail if any HDFS encryption zone were specified without choosing Hadoop as an application.
- Changed the default HDFS write policy from `RoundRobin` to `AvailableSpaceVolumeChoosingPolicy`. Some volumes were not properly utilized with the RoundRobin configuration, which resulted in failed core nodes and an unreliable HDFS.
- Fixed an issue with the EMRFS CLI, which would cause an exception when creating the default DynamoDB metadata table for consistent views.
- Fixed a deadlock issue in EMRFS that potentially occurred during multipart rename and copy operations.
- Fixed an issue with EMRFS that caused the CopyPart size default to be 5 MB. The default is now properly set at 128 MB.
- Fixed an issue with the Zeppelin upstart configuration that potentially prevented you from stopping the service.

- Fixed an issue with Spark and Zeppelin, which prevented you from using the `s3a://` URI scheme because `/usr/lib/hadoop/hadoop-aws.jar` was not properly loaded in their respective classpath.
- Backported HUE-2484.
- Backported a commit from Hue 3.9.0 (no JIRA exists) to fix an issue with the HBase browser sample.
- Backported HIVE-9073.

Release 4.6.0

Release date: April 21, 2016

Features

The following features are available in this release:

- Added HBase 1.2.0
- Added Zookeeper-Sandbox 3.4.8
- Upgraded to Presto-Sandbox 0.143
- Amazon EMR releases are now based on Amazon Linux 2016.03.0. For more information, see https://aws.amazon.com/amazon-linux-ami/2016.03-release-notes/.

Issue affecting Throughput Optimized HDD (st1) EBS volume types

An issue in the Linux kernel versions 4.2 and above significantly affects performance on Throughput Optimized HDD (st1) EBS volumes for EMR. This release (emr-4.6.0) uses kernel version 4.4.5 and hence is impacted. Therefore, we recommend not using emr-4.6.0 if you want to use st1 EBS volumes. You can use emr-4.5.0 or prior Amazon EMR releases with st1 without impact. In addition, we provide the fix with future releases.

Python defaults

Python 3.4 is now installed by default, but Python 2.7 remains the system default. You may configure Python 3.4 as the system default using either a bootstrap action; you can use the configuration API to set PYSPARK_PYTHON export to `/usr/bin/python3.4` in the `spark-env` classification to affect the Python version used by PySpark.

Java 8

Except for Presto, OpenJDK 1.7 is the default JDK used for all applications. However, both OpenJDK 1.7 and 1.8 are installed. For information about how to set JAVA_HOME for applications, see Configuring applications to use Java 8.

Known issues resolved from previous releases

- Fixed an issue where application provisioning would sometimes randomly fail due to a generated password.
- Previously, `mysqld` was installed on all nodes. Now, it is only installed on the master instance and only if the chosen application includes `mysql-server` as a component. Currently, the following applications include the `mysql-server` component: HCatalog, Hive, Hue, Presto-Sandbox, and Sqoop-Sandbox.
- Changed `yarn.scheduler.maximum-allocation-vcores` to 80 from the default of 32, which fixes an issue introduced in emr-4.4.0 that mainly occurs with Spark while using the `maximizeResourceAllocation` option in a cluster whose core instance type is one of a few large instance types that have the YARN vcores set higher than 32; namely c4.8xlarge, cc2.8xlarge, hs1.8xlarge, i2.8xlarge, m2.4xlarge, r3.8xlarge, d2.8xlarge, or m4.10xlarge were affected by this issue.
- s3-dist-cp now uses EMRFS for all Amazon S3 nominations and no longer stages to a temporary HDFS directory.
- Fixed an issue with exception handling for client-side encryption multipart uploads.
- Added an option to allow users to change the Amazon S3 storage class. By default this setting is `STANDARD`. The `emrfs-site` configuration classification setting is `fs.s3.storageClass` and the possible values are `STANDARD`, `STANDARD_IA`, and `REDUCED_REDUNDANCY`. For more information about storage classes, see Storage classes in the Amazon Simple Storage Service User Guide.

Release 4.5.0

Release date: April 4, 2016

Features

The following features are available in this release:

- Upgraded to Spark 1.6.1
- Upgraded to Hadoop 2.7.2
- Upgraded to Presto 0.140
- Added AWS KMS support for Amazon S3 server-side encryption.

Known issues resolved from previous releases

- Fixed an issue where MySQL and Apache servers would not start after a node was rebooted.
- Fixed an issue where IMPORT did not work correctly with non-partitioned tables stored in Amazon S3
- Fixed an issue with Presto where it requires the staging directory to be `/mnt/tmp` rather than `/tmp` when writing to Hive tables.

Release 4.4.0

Release date: March 14, 2016

Features

The following features are available in this release:

- Added HCatalog 1.0.0
- Added Sqoop-Sandbox 1.4.6
- Upgraded to Presto 0.136
- Upgraded to Zeppelin 0.5.6
- Upgraded to Mahout 0.11.1
- Enabled `dynamicResourceAllocation` by default.
- Added a table of all configuration classifications for the release. For more information, see the Configuration Classifications table in Configuring applications.

Known issues resolved from previous releases

- Fixed an issue where the `maximizeResourceAllocation` setting would not reserve enough memory for YARN ApplicationMaster daemons.

- Fixed an issue encountered with a custom DNS. If any entries in `resolve.conf` precede the custom entries provided, then the custom entries are not resolvable. This behavior was affected by clusters in a VPC where the default VPC name server is inserted as the top entry in `resolve.conf`.
- Fixed an issue where the default Python moved to version 2.7 and boto was not installed for that version.
- Fixed an issue where YARN containers and Spark applications would generate a unique Ganglia round robin database (rrd) file, which resulted in the first disk attached to the instance filling up. Because of this fix, YARN container level metrics have been disabled and Spark application level metrics have been disabled.
- Fixed an issue in log pusher where it would delete all empty log folders. The effect was that the Hive CLI was not able to log because log pusher was removing the empty `user` folder under `/var/log/hive`.
- Fixed an issue affecting Hive imports, which affected partitioning and resulted in an error during import.
- Fixed an issue where EMRFS and s3-dist-cp did not properly handle bucket names that contain periods.
- Changed a behavior in EMRFS so that in versioning-enabled buckets the `_$folder$` marker file is not continuously created, which may contribute to improved performance for versioning-enabled buckets.
- Changed the behavior in EMRFS such that it does not use instruction files except for cases where client-side encryption is enabled. If you want to delete instruction files while using client-side encryption, you can set the emrfs-site.xml property, `fs.s3.cse.cryptoStorageMode.deleteInstructionFiles.enabled`, to true.
- Changed YARN log aggregation to retain logs at the aggregation destination for two days. The default destination is your cluster's HDFS storage. If you want to change this duration, change the value of `yarn.log-aggregation.retain-seconds` using the `yarn-site` configuration classification when you create your cluster. As always, you can save your application logs to Amazon S3 using the `log-uri` parameter when you create your cluster.

Patches applied

The following patches from open source projects were included in this release:

- HIVE-9655

- HIVE-9183
- HADOOP-12810

Release 4.3.0

Release date: January 19, 2016

Features

The following features are available in this release:

- Upgraded to Hadoop 2.7.1
- Upgraded to Spark 1.6.0
- Upgraded Ganglia to 3.7.2
- Upgraded Presto to 0.130

Amazon EMR made some changes to `spark.dynamicAllocation.enabled` when it is set to true; it is false by default. When set to true, this affects the defaults set by the `maximizeResourceAllocation` setting:

- If `spark.dynamicAllocation.enabled` is set to true, `spark.executor.instances` is not set by `maximizeResourceAllocation`.
- The `spark.driver.memory` setting is now configured based on the instance types in the cluster in a similar way to how `spark.executors.memory` is set. However, because the Spark driver application may run on either the master or one of the core instances (for example, in YARN client and cluster modes, respectively), the `spark.driver.memory` setting is set based on the instance type of the smaller instance type between these two instance groups.
- The `spark.default.parallelism` setting is now set at twice the number of CPU cores available for YARN containers. In previous releases, this was half that value.
- The calculations for the memory overhead reserved for Spark YARN processes was adjusted to be more accurate, resulting in a small increase in the total amount of memory available to Spark (that is, `spark.executor.memory`).

Known issues resolved from the previous releases

- YARN log aggregation is now enabled by default.

- Fixed an issue where logs would not be pushed to a cluster's Amazon S3 logs bucket when YARN log aggregation was enabled.
- YARN container sizes now have a new minimum of 32 across all node types.
- Fixed an issue with Ganglia that caused excessive disk I/O on the master node in large clusters.
- Fixed an issue that prevented applications logs from being pushed to Amazon S3 when a cluster is shutting down.
- Fixed an issue in EMRFS CLI that caused certain commands to fail.
- Fixed an issue with Zeppelin that prevented dependencies from being loaded in the underlying SparkContext.
- Fixed an issue that resulted from issuing a resize attempting to add instances.
- Fixed an issue in Hive where CREATE TABLE AS SELECT makes excessive list calls to Amazon S3.
- Fixed an issue where large clusters would not provision properly when Hue, Oozie, and Ganglia are installed.
- Fixed an issue in s3-dist-cp where it would return a zero exit code even if it failed with an error.

Patches applied

The following patches from open source projects were included in this release:

- OOZIE-2402
- HIVE-12502
- HIVE-10631
- HIVE-12213
- HIVE-10559
- HIVE-12715
- HIVE-10685

Release 4.2.0

Release date: November 18, 2015

Features

The following features are available in this release:

- Added Ganglia support
- Upgraded to Spark 1.5.2
- Upgraded to Presto 0.125
- Upgraded Oozie to 4.2.0
- Upgraded Zeppelin to 0.5.5
- Upgraded the AWS SDK for Java to 1.10.27

Known issues resolved from the previous releases

- Fixed an issue with the EMRFS CLI where it did not use the default metadata table name.
- Fixed an issue encountered when using ORC-backed tables in Amazon S3.
- Fixed an issue encountered with a Python version mismatch in the Spark configuration.
- Fixed an issue when a YARN node status fails to report because of DNS issues for clusters in a VPC.
- Fixed an issue encountered when YARN decommissioned nodes, resulting in hanged applications or the inability to schedule new applications.
- Fixed an issue encountered when clusters terminated with status TIMED_OUT_STARTING.
- Fixed an issue encountered when including the EMRFS Scala dependency in other builds. The Scala dependency has been removed.

Configure applications

To override the default configurations for an application, you can supply a configuration object. You can either use a shorthand syntax to provide the configuration, or you can reference the configuration object in a JSON file. Configuration objects consist of a classification, properties, and optional nested configurations. Properties correspond to the application settings you want to change. You can specify multiple classifications for multiple applications in a single JSON object.

> **Warning**
>
> Amazon EMR Describe and List API operations emit custom and configurable settings, which are used as a part of Amazon EMR job flows, in plaintext. To provide sensitive information, such as passwords, in these settings, see Store sensitive configuration data in AWS Secrets Manager.

The configuration classifications that are available vary by Amazon EMR release version. For a list of configuration classifications that are supported in a particular release version, refer to the page for that release version under About Amazon EMR Releases.

The following is example JSON file for a list of configurations.

```
[
  {
    "Classification": "core-site",
    "Properties": {
      "hadoop.security.groups.cache.secs": "250"
    }
  },
  {
    "Classification": "mapred-site",
    "Properties": {
      "mapred.tasktracker.map.tasks.maximum": "2",
      "mapreduce.map.sort.spill.percent": "0.90",
      "mapreduce.tasktracker.reduce.tasks.maximum": "5"
    }
  }
]
```

A configuration classification often maps to an application-specific configuration file. For example, the `hive-site` classification maps to settings in the `hive-site.xml` configuration file for Hive. An exception to this is the no longer supported bootstrap action `configure-daemons`, which is used to set environment parameters such as `--namenode-heap-size`. Options like this are subsumed into the `hadoop-env` and `yarn-env` classifications with their own nested export classifications. If any classification ends in `env`, use the export sub-classification.

Another exception is `s3get`, which is used to place a customer `EncryptionMaterialsProvider` object on each node in a cluster for use in client-side encryption. An option was added to the `emrfs-site` classification for this purpose.

The following is an example of the `hadoop-env` classification.

```
[
  {
    "Classification": "hadoop-env",
    "Properties": {

    },
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "HADOOP_DATANODE_HEAPSIZE": "2048",
          "HADOOP_NAMENODE_OPTS": "-XX:GCTimeRatio=19"
        },
        "Configurations": [

        ]
      }
    ]
  }
]
```

The following is an example of the yarn-env classification.

```
[
  {
    "Classification": "yarn-env",
    "Properties": {
```

```
    },
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "YARN_RESOURCEMANAGER_OPTS": "-Xdebug -Xrunjdwp:transport=dt_socket"
        },
        "Configurations": [

        ]
      }
    ]
  }
]
```

The following settings do not belong to a configuration file but are used by Amazon EMR to potentially configure multiple settings on your behalf.

Settings curated by Amazon EMR

Application	Release label classification	Valid properties	When to use
Spark	`spark`	`maximizeResourceAllocation`	Configure executors to utilize the maximum resources of each node.

Topics

- Configure applications when you create a cluster
- Reconfigure an instance group in a running cluster
- Store sensitive configuration data in AWS Secrets Manager
- Configure applications to use a specific Java Virtual Machine

Configure applications when you create a cluster

When you create a cluster, you can override the default configurations for applications using the Amazon EMR console, the AWS Command Line Interface (AWS CLI), or the AWS SDK.

To override the default configuration for an application, you specify custom values in a configuration classification. A configuration classification corresponds to a configuration XML file for an application, such as `hive-site.xml`.

Configuration classifications vary by Amazon EMR release version. For a list of configuration classifications that are available in a specific release version, see the release detail page. For example, [Amazon EMR release 6.4.0.](#)

Supply a configuration in the console when you create a cluster

To supply a configuration, navigate to the **Create cluster** page and expand **Software settings**. You can then enter the configuration directly by using either JSON or a shorthand syntax demonstrated in shadow text in the console. Otherwise, you can provide an Amazon S3 URI for a file with a JSON `Configurations` object.

To supply a configuration for an instance group, choose a cluster in your list of clusters, then choose the **Configurations** tab. In the **Instance group configurations** table, choose the instance group to edit, then choose **Reconfigure**.

Supply a configuration using the AWS CLI when you create a cluster

You can provide a configuration to **create-cluster** by supplying a path to a JSON file stored locally or in Amazon S3. The following example assumes that you are using default roles for Amazon EMR and that the roles have been created. If you need to create the roles, run `aws emr create-default-roles` first.

If your configuration is in your local directory, you can use the following example command.

```
aws emr create-cluster --use-default-roles --release-label emr-7.2.0 --applications
 Name=Hive \
--instance-type m5.xlarge --instance-count 3 --configurations file://./
configurations.json
```

If your configuration is in an Amazon S3 path, you'll need to set up the following workaround before passing the Amazon S3 path to the `create-cluster` command.

```
#!/bin/sh
# Assume the ConfigurationS3Path is not public, and its present in the same AWS account
 as the EMR cluster
ConfigurationS3Path="s3://my-bucket/config.json"
# Get a presigned HTTP URL for the s3Path
```

```
ConfigurationURL=`aws s3 presign $ConfigurationS3Path --expires-in 300`
# Fetch the presigned URL, and minify the JSON so that it spans only a single line
Configurations=`curl $ConfigurationURL | jq -c .`
aws emr create-cluster --use-default-roles --release-label emr-5.34.0 --instance-type
 m5.xlarge --instance-count 2 --applications Name=Hadoop Name=Spark --configurations
 $Configurations
```

Supply a configuration using the Java SDK when you create a cluster

The following program excerpt shows how to supply a configuration using the AWS SDK for Java.

```
Application hive = new Application().withName("Hive");

Map<String,String> hiveProperties = new HashMap<String,String>();
 hiveProperties.put("hive.join.emit.interval","1000");
 hiveProperties.put("hive.merge.mapfiles","true");

Configuration myHiveConfig = new Configuration()
 .withClassification("hive-site")
 .withProperties(hiveProperties);

RunJobFlowRequest request = new RunJobFlowRequest()
 .withName("Create cluster with ReleaseLabel")
 .withReleaseLabel("emr-5.20.0")
 .withApplications(hive)
 .withConfigurations(myHiveConfig)
 .withServiceRole("EMR_DefaultRole")
 .withJobFlowRole("EMR_EC2_DefaultRole")
 .withInstances(new JobFlowInstancesConfig()
  .withEc2KeyName("myEc2Key")
  .withInstanceCount(3)
  .withKeepJobFlowAliveWhenNoSteps(true)
  .withMasterInstanceType("m4.large")
  .withSlaveInstanceType("m4.large")
 );
```

Reconfigure an instance group in a running cluster

With Amazon EMR version 5.21.0 and later, you can reconfigure cluster applications and specify additional configuration classifications for each instance group in a running cluster. To do so, you can use the Amazon EMR console, the AWS Command Line Interface (AWS CLI), or the AWS SDK.

When you update an application configuration for an instance group in the new Amazon EMR console, the console attempts to merge the new configuration with the existing configuration to create a new, active configuration. In the unusual case where Amazon EMR can't merge the configuration, the console alerts you.

After you submit a reconfiguration request for an instance group, Amazon EMR assigns a version number to the new configuration specification. You can track the version number of a configuration, or the state of an instance group, by viewing the CloudWatch events. For more information, see Monitor CloudWatch Events.

> **Note**
>
> You can only override, and not delete, cluster configurations that were specified during cluster creation. If there are differences between the existing configuration and the file that you supply, Amazon EMR resets manually modified configurations, such as configurations that you have modified while connected to your cluster using SSH, to the cluster defaults for the specified instance group.

Considerations when you reconfigure an instance group

Reconfiguration actions

When you submit a reconfiguration request using the Amazon EMR console, the AWS Command Line Interface (AWS CLI), or the AWS SDK, Amazon EMR checks the existing on-cluster configuration file. If there are differences between the existing configuration and the file that you supply, Amazon EMR initiates reconfiguration actions, restarts some applications, and resets any manually modified configurations, such as configurations that you have modified while connected to your cluster using SSH, to the cluster defaults for the specified instance group.

> **Note**
>
> Amazon EMR performs some default actions during every instance group reconfiguration. These default actions might conflict with cluster customizations that you have made, and result in reconfiguration failures. For information about how to troubleshoot reconfiguration failures, see Troubleshoot instance group reconfiguration.

Amazon EMR also initiates reconfiguration actions for the configuration classifications that you specify in your request. For a complete list of these actions, see the Configuration Classifications section for the version of Amazon EMR that you use. For example, 6.2.0 Configuration Classifications.

> **Note**
>
> The Amazon EMR Release Guide only lists reconfiguration actions starting with Amazon EMR versions 5.32.0 and 6.2.0.

Service disruption

Amazon EMR follows a rolling process to reconfigure instances in the Task and Core instance groups. Only 10 percent of the instances in an instance group are modified and restarted at a time. This process takes longer to finish but reduces the chance of potential application failure in a running cluster.

To run YARN jobs during a YARN restart, you can either create an Amazon EMR cluster with multiple master nodes or set `yarn.resourcemanager.recovery.enabled` to `true` in your `yarn-site` configuration classification. For more information about using multiple master nodes, see High availability YARN ResourceManager.

Application validation

Amazon EMR checks that each application on the cluster is running after the reconfiguration restart process. If any application is unavailable, the overall reconfiguration operation fails. If a reconfiguration operation fails, Amazon EMR reverses the configuration parameters to the previous working version.

> **Note**
>
> To avoid reconfiguration failure, we recommend that you only install applications on your cluster that you plan to use. We also recommend that you make sure all cluster applications are healthy and running before you submit a reconfiguration request.

Types of reconfiguration

You can reconfigure an instance group in one of two ways:

- **Overwrite**. Default reconfiguration method and the only one available in Amazon EMR releases earlier than 5.35.0 and 6.6.0. This reconfiguration method indiscriminately overwrites any on-cluster files with the newly submitted configuration set. The method erases any changes to configuration files made outside the reconfiguration API.
- **Merge**. Reconfiguration method supported for Amazon EMR releases 5.35.0 and 6.6.0 and later, except from the Amazon EMR console, where no version supports it. This reconfiguration method merges the newly submitted configurations with configurations that already exist on the cluster. This option only adds or modifies the new configurations that you submit. It preserves existing configurations.

> **Note**
>
> Amazon EMR continues to overwrite some essential Hadoop configurations that it needs to ensure that the service is running correctly.

Limitations

When you reconfigure an instance group in a running cluster, consider the following limitations:

- Non-YARN applications can fail during restart or cause cluster issues, especially if the applications aren't configured properly. Clusters approaching maximum memory and CPU usage may run into issues after the restart process. This is especially true for the master instance group.
- You can't submit a reconfiguration request when an instance group is being resized. If a reconfiguration is initiated while an instance group is resizing, reconfiguration cannot start until the instance group has completed resizing, and vice versa.
- After reconfiguring an instance group, Amazon EMR restarts the applications to allow the new configurations to take effect. Job failure or other unexpected application behavior might occur if the applications are in use during reconfiguration.
- If a reconfiguration for an instance group fails, Amazon EMR reverses the configuration parameters to the previous working version. If the reversion process fails too, you must submit a new `ModifyInstanceGroup` request to recover the instance group from the `SUSPENDED` state.
- Reconfiguration requests for Phoenix configuration classifications are only supported in Amazon EMR version 5.23.0 and later, and are not supported in Amazon EMR version 5.21.0 or 5.22.0.
- Reconfiguration requests for HBase configuration classifications are only supported in Amazon EMR version 5.30.0 and later, and are not supported in Amazon EMR versions 5.23.0 through 5.29.0.

- Amazon EMR supports application reconfiguration requests on an Amazon EMR cluster with multiple primary nodes only in Amazon EMR versions 5.27.0 and later.
- Reconfiguring `hdfs-encryption-zones` classification or any of the Hadoop KMS configuration classifications is not supported on an Amazon EMR cluster with multiple primary nodes.
- Amazon EMR currently doesn't support certain reconfiguration requests for the capacity scheduler that require restarting the YARN ResourceManager. For example, you cannot completely remove a queue.

Reconfigure an instance group in the console

Note

The Amazon EMR console does not support **Merge** type reconfigurations.

1. Open the Amazon EMR console at https://console.aws.amazon.com/emr
2. In the cluster list under **Name**, choose the active cluster that you want to reconfigure.
3. Open the cluster details page for the cluster, and go to the **Configurations** tab.
4. In the **Filter** drop-down list, select the instance group that you want to reconfigure.
5. In the **Reconfigure** drop-down menu, choose either **Edit in table** or **Edit in JSON file**.
   - **Edit in table** - In the configuration classification table, edit the property and value for existing configurations, or choose **Add configuration** to supply additional configuration classifications.
   - **Edit in JSON file** - Enter the configuration directly in JSON, or use shorthand syntax (demonstrated in shadow text). Otherwise, provide an Amazon S3 URI for a file with a JSON `Configurations` object.

   Note

   The **Source** column in the configuration classification table indicates whether the configuration is supplied when you create a cluster, or when you specify additional configurations for this instance group. You can edit the configurations for an instance group from both sources. You cannot delete initial cluster configurations, but you can override them for an instance group.

You can also add or edit nested configuration classifications directly in the table. For example, to supply an additional `export` sub-classification of `hadoop-env`, add a `hadoop.export` configuration classification in the table. Then, provide a specific property and value for this classification.

6. (Optional) Select **Apply this configuration to all active instance groups**.
7. Save the changes.

Reconfigure an instance group using the CLI

Use the **modify-instance-groups** command to specify a new configuration for an instance group in a running cluster.

> **Note**
>
> In the following examples, replace *<j-2AL4XXXXXX5T9>* with your cluster ID, and replace *<ig-1xxxxxxx9>* with your instance group ID.

Example – Replace a configuration for an instance group

The following example references a configuration JSON file called `instanceGroups.json` to edit the property of the YARN NodeManager disk health checker for an instance group.

1. Prepare your configuration classification, and save it as `instanceGroups.json` in the same directory where you will run the command.

```
[
   {
      "InstanceGroupId":"<ig-1xxxxxxx9>",
      "Configurations":[
         {
            "Classification":"yarn-site",
            "Properties":{
               "yarn.nodemanager.disk-health-checker.enable":"true",
               "yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-
percentage":"100.0"
            },
```

```
            "Configurations":[]
         }
      ]
   }
]
```

2. Run the following command.

```
aws emr modify-instance-groups --cluster-id <j-2AL4XXXXXX5T9> \
--instance-groups file://instanceGroups.json
```

Example – Add a configuration to an instance group

If you want to add a configuration to an instance group, you must include all previously specified configurations for that instance group in your new `ModifyInstanceGroup` request. Otherwise, the previously specified configurations are removed.

The following example adds a property for the YARN NodeManager virtual memory checker. The configuration also includes previously specified values for the YARN NodeManager disk health checker so that the values won't be overwritten.

1. Prepare the following contents in `instanceGroups.json` and save it in the same directory where you will run the command.

```
[
   {
      "InstanceGroupId":"<ig-1xxxxxxx9>",
      "Configurations":[
         {
            "Classification":"yarn-site",
            "Properties":{
               "yarn.nodemanager.disk-health-checker.enable":"true",
               "yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-
percentage":"100.0",
               "yarn.nodemanager.vmem-check-enabled":"true",
               "yarn.nodemanager.vmem-pmem-ratio":"3.0"
            },
            "Configurations":[]
         }
```

```
        ]
    }
]
```

2. Run the following command.

```
aws emr modify-instance-groups --cluster-id <j-2AL4XXXXXX5T9> \
--instance-groups file://instanceGroups.json
```

Example – Add a configuration to an instance group with Merge type reconfiguration

When you want to use the default **Overwrite** reconfiguration method to add a configuration, you must include all previously specified configurations for that instance group in your new `ModifyInstanceGroup` request. Otherwise, the **Overwrite** removes the configurations that you previously specified. You don't need to do this with **Merge** reconfiguration. Instead, you must ensure that your request only includes the new configurations are included.

The following example adds a property for the YARN NodeManager virtual memory checker. Because this is a **Merge** type reconfiguration, it does not overwrite previously specified values for the YARN NodeManager disk health checker.

1. Prepare the following contents in `instanceGroups.json` and save it in the same directory where you will run the command.

```
[
   {"InstanceGroupId":"<ig-1xxxxxxx9>",
    "ReconfigurationType" :"MERGE",
      "Configurations":[
         {"Classification":"yarn-site",
            "Properties":{
               "yarn.nodemanager.vmem-check-enabled":"true",
               "yarn.nodemanager.vmem-pmem-ratio":"3.0"
            },
            "Configurations":[]
         }
      ]
   }
]
```

2. Run the following command.

```
aws emr modify-instance-groups --cluster-id <j-2AL4XXXXXX5T9> \
--instance-groups file://instanceGroups.json
```

Example – Delete a configuration for an instance group

To delete a configuration for an instance group, submit a new reconfiguration request that excludes the previous configuration.

> **Note**
>
> You can only override the initial *cluster* configuration. You cannot delete it.

For example, to delete the configuration for the YARN NodeManager disk health checker from the previous example, submit a new `instanceGroups.json` with the following contents.

```
[
   {
      "InstanceGroupId":"<ig-1xxxxxxx9>",
      "Configurations":[
         {
            "Classification":"yarn-site",
            "Properties":{
               "yarn.nodemanager.vmem-check-enabled":"true",
               "yarn.nodemanager.vmem-pmem-ratio":"3.0"
            },
            "Configurations":[]
         }
      ]
   }
]
```

Note

To delete all of the configurations in your last reconfiguration request, submit a reconfiguration request with an empty array of configurations. For example,

```
[
   {
      "InstanceGroupId":"<ig-1xxxxxxx9>",
      "Configurations":[]
   }
]
```

Example – Reconfigure and resize an instance group in one request

The following example JSON demonstrates how to reconfigure and resize an instance group in the same request.

```
[
   {
      "InstanceGroupId":"<ig-1xxxxxxx9>",
      "InstanceCount":5,
      "EC2InstanceIdsToTerminate":["i-123"],
      "ForceShutdown":true,
      "ShrinkPolicy":{
         "DecommissionTimeout":10,
         "InstanceResizePolicy":{
            "InstancesToTerminate":["i-123"],
            "InstancesToProtect":["i-345"],
            "InstanceTerminationTimeout":20
         }
      },
      "Configurations":[
         {
            "Classification":"yarn-site",
            "Configurations":[],
            "Properties":{
               "yarn.nodemanager.disk-health-checker.enable":"true",
               "yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-
percentage":"100.0"
            }
```

```
        }
      ]
    }
]
```

Reconfigure an instance group using the Java SDK

Note

In the following examples, replace *<j-2AL4XXXXXX5T9>* with your cluster ID, and replace *<ig-1xxxxxxx9>* with your instance group ID.

The following code snippet provides a new configuration for an instance group using the AWS SDK for Java.

```
AWSCredentials credentials = new BasicAWSCredentials("access-key", "secret-key");
AmazonElasticMapReduce emr = new AmazonElasticMapReduceClient(credentials);

Map<String,String> hiveProperties = new HashMap<String,String>();
hiveProperties.put("hive.join.emit.interval","1000");
hiveProperties.put("hive.merge.mapfiles","true");

Configuration configuration = new Configuration()
    .withClassification("hive-site")
    .withProperties(hiveProperties);

InstanceGroupModifyConfig igConfig = new InstanceGroupModifyConfig()
    .withInstanceGroupId("<ig-1xxxxxxx9>")
    .withReconfigurationType("MERGE");
    .withConfigurations(configuration);

ModifyInstanceGroupsRequest migRequest = new ModifyInstanceGroupsRequest()
    .withClusterId("<j-2AL4XXXXXX5T9>")
    .withInstanceGroups(igConfig);

emr.modifyInstanceGroups(migRequest);
```

The following code snippet deletes a previously specified configuration for an instance group by supplying an empty array of configurations.

```
List<Configuration> configurations = new ArrayList<Configuration>();

InstanceGroupModifyConfig igConfig = new InstanceGroupModifyConfig()
    .withInstanceGroupId("<ig-1xxxxxxx9>")
    .withConfigurations(configurations);

ModifyInstanceGroupsRequest migRequest = new ModifyInstanceGroupsRequest()
    .withClusterId("<j-2AL4XXXXXX5T9>")
    .withInstanceGroups(igConfig);

emr.modifyInstanceGroups(migRequest);
```

Troubleshoot instance group reconfiguration

If the reconfiguration process for an instance group fails, Amazon EMR reverts the reconfiguration and logs a failure message using an Amazon CloudWatch event. The event provides a brief summary of the reconfiguration failure. It lists the instances for which reconfiguration has failed and corresponding failure messages. The following is an example failure message.

```
The reconfiguration operation for instance group ig-1xxxxxxx9 in Amazon EMR
 cluster j-2AL4XXXXXX5T9 (ExampleClusterName)
failed at 2021-01-01 00:00 UTC and took 2 minutes to fail. Failed configuration version
 is example12345.
Failure message: Instance i-xxxxxxx1, i-xxxxxxx2, i-xxxxxxx3 failed with message "This
 is an example failure message".
```

To gather more data about a reconfiguration failure, you can check the node provisioning logs. Doing so is particularly useful when you receive a message like the following.

```
i-xxxxxxx1 failed with message “Unable to complete transaction and some changes were
 applied.”
```

On the node

To access node provisioning logs by connecting to a node

1. Use SSH to connect to the node on which reconfiguration has failed. For instructions, see Connect to your Linux instance in the *Amazon EC2 User Guide for Linux Instances*.
2. Navigate to the following directory, which contains the node provisioning log files.

```
/mnt/var/log/provision-node/
```

3. Open the `reports` subdirectory and search for the node provisioning report for your reconfiguration. The `reports` directory organizes logs by reconfiguration version number, universally unique identifier (UUID), Amazon EC2 instance IP address, and timestamp. Each report is a compressed YAML file that contains detailed information about the reconfiguration process.

   The following is an example report file name and path.

```
/reports/2/ca598xxx-cxxx-4xxx-bxxx-6dbxxxxxxxxx/ip-10-73-xxx-
xxx.ec2.internal/202104061715.yaml.gz
```

4. You can examine a report using a file viewer like `zless`, as in the following example.

```
zless 202104061715.yaml.gz
```

Amazon S3

To access node provisioning logs using Amazon S3

1. Sign in to the AWS Management Console and open the Amazon S3 console at https://console.aws.amazon.com/s3/.
2. Open the Amazon S3 bucket that you specified when you configured the cluster to archive log files.
3. Navigate to the following folder, which contains the node provisioning log files:

```
DOC-EXAMPLE-BUCKET/elasticmapreduce/<cluster id>/node/<instance id>/provision-
node/
```

4. Open the `reports` folder and search for the node provisioning report for your reconfiguration. The `reports` folder organizes logs by reconfiguration version number, universally unique identifier (UUID), Amazon EC2 instance IP address, and timestamp. Each report is a compressed YAML file that contains detailed information about the reconfiguration process.

   The following is an example report file name and path.

   ```
   /reports/2/ca598xxx-cxxx-4xxx-bxxx-6dbxxxxxxxxx/ip-10-73-xxx-
   xxx.ec2.internal/202104061715.yaml.gz
   ```

5. To view a log file, you can download it from Amazon S3 to your local machine as a text file. For instructions, see Downloading an object.

Each log file contains a detailed provisioning report for the associated reconfiguration. To find error message information, you can search for the `err` log level of a report. Report format depends on the version of Amazon EMR on your cluster.

The following example shows error information for Amazon EMR release versions earlier than 5.32.0 and 6.2.0.

```
- !ruby/object:Puppet::Util::Log
      level: !ruby/sym err
      tags:
        - err
      message: "Example detailed error message."
      source: Puppet
      time: 2021-01-01 00:00:00.000000 +00:00
```

Amazon EMR release versions 5.32.0 and 6.2.0 and later use the following format instead.

```
- level: err
  message: 'Example detailed error message.'
  source: Puppet
  tags:
  - err
  time: '2021-01-01 00:00:00.000000 +00:00'
  file:
  line:
```

Store sensitive configuration data in AWS Secrets Manager

The Amazon EMR describe and list API operations that emit custom configuration data (such as `DescribeCluster` and `ListInstanceGroups`) do so in plaintext. Amazon EMR integrates with AWS Secrets Manager so that you can store your data in Secrets Manager and use the secret ARN in your configurations. This way, you don't pass sensitive configuration data to Amazon EMR in plaintext and expose it to external APIs. If you indicate that a key-value pair contains an ARN for a secret stored in Secrets Manager, Amazon EMR retrieves this secret when it sends configuration data to the cluster. Amazon EMR doesn't send the annotation when it uses external APIs to display the configuration.

Create a secret

To create a secret, follow the steps in Create an AWS Secrets Manager secret in the *AWS Secrets Manager User Guide*. In **Step 3**, you must choose the **Plaintext** field to enter your sensitive value.

Note that while Secrets Manager allows a secret to contain up to 65536 bytes, Amazon EMR limits the combined length of the property key (excluding the annotation) and the retrieved secret value to 1024 characters.

Grant Amazon EMR access to retrieve the secret

Amazon EMR uses an IAM service role to provision and manage clusters for you. The service role for Amazon EMR defines the allowable actions for Amazon EMR when it provisions resources and performs service-level tasks that aren't performed in the context of an Amazon EC2 instance running within a cluster. For more information about service roles, see Service role for Amazon EMR (EMR role) and Customize IAM roles.

To allow Amazon EMR to retrieve the secret value from Secrets Manager, add the following policy statement to your Amazon EMR role when you launch your cluster.

```
{
   "Sid":"AllowSecretsRetrieval",
   "Effect":"Allow",
   "Action":"secretsmanager:GetSecretValue",
   "Resource":[
      "arn:aws:secretsmanager:<region>:<aws-account-id>:secret:<secret-name>"
   ]
}
```

If you create the secret with a customer-managed AWS KMS key, you must also add `kms:Decrypt` permission to the Amazon EMR role for the key that you use. For more information, see [Authentication and access control for AWS Secrets Manager](#) in the *AWS Secrets Manager User Guide*.

Use the secret in a configuration classification

You can add the `EMR.secret@` annotation to any configuration property to indicate that its key-value pair contains an ARN for a secret stored in Secrets Manager.

The following example shows how to provide a secret ARN in a configuration classification:

```
{
   "Classification":"core-site",
   "Properties":{
      "presto.s3.access-key":"<sensitive-access-key>",
      "EMR.secret@presto.s3.secret-key":"arn:aws:secretsmanager:<region>:<aws-account-id>:secret:<secret-name>"
   }
}
```

When you create your cluster and submit your annotated configuration, Amazon EMR validates the configuration properties. If your configuration is valid, Amazon EMR strips the annotation from the configuration and retrieves the secret from Secrets Manager to create the actual configuration before applying it to the cluster:

```
{
   "Classification":"core-site",
   "Properties":{
      "presto.s3.access-key":"<sensitive-access-key>",
      "presto.s3.secret-key":"<my-secret-key-retrieved-from-Secrets-Manager>"
   }
}
```

When you call an action like `DescribeCluster`, Amazon EMR returns the current application configuration on the cluster. If an application configuration property is marked as containing a secret ARN, then the application configuration returned by the `DescribeCluster` call contains the ARN and not the secret value. This ensures that the secret value is only visible on the cluster:

```
{
   "Classification":"core-site",
   "Properties":{
```

```
      "presto.s3.access-key":"<sensitive-access-key>",
      "presto.s3.secret-key":"arn:aws:secretsmanager:<region>:<aws-account-
id>:secret:<secret-name>"
    }
}
```

Update the secret value

Amazon EMR retrieves the secret value from an annotated configuration whenever the attached instance group is starting, reconfiguring, or resizing. You can use Secrets Manager to modify the value of a secret used in the configuration of a running cluster. When you do, you can submit a reconfiguration request to each instance group that you want to receive the updated value. For more information on how to reconfigure an instance group, and things to consider when you do it, see Reconfigure an instance group in a running cluster.

Configure applications to use a specific Java Virtual Machine

Amazon EMR releases have different default Java Virtual Machine (JVM) versions. This page explains the JVM support for different releases and applications.

Considerations

For information about the supported Java versions for applications, see the application pages in the Amazon EMR Release Guide.

- Amazon EMR only supports running one runtime version in a cluster, and doesn't support running different nodes or applications on different runtime versions on the same cluster.
- For Amazon EMR 7.x, the default Java Virtual Machine (JVM) is Java 17 for applications that support Java 17, with the exception of Apache Livy. For more information about the supported JDK versions for applications, see the corresponding release page in the Amazon EMR Release Guide.
- Starting with Amazon EMR 7.1.0, Flink supports and is set to Java 17 by default. To use a different version Java runtime, override the settings in `flink-conf`. For more information about configuring Flink to use Java 8 or Java 11, see Configure Flink to run with Java 11.
- For Amazon EMR 5.x and 6.x, the default Java Virtual Machine (JVM) is Java 8.
  - For Amazon EMR releases 6.12.0 and higher, some applications also support Java 11 and 17.
  - For Amazon EMR releases 6.9.0 and higher, Trino supports Java 17 as default. For more information about Java 17 with Trino, see Trino updates to Java 17 on the Trino blog.

Keep in mind the following application-specific considerations when you choose your runtime version:

Application-specific Java configuration notes

Application	Java configuration notes
Spark	To run Spark with a non-default Java version, you must configure both Spark and Hadoop. For examples, see [Override the JVM](#). • Configure `JAVA_HOME` in `spark-env` to update the Java runtime of primary instance processes. For example, spark-sub mit, spark-shell, and Spark History Server. • Modify the Hadoop configuration to update the Java runtime of the Spark executors and the YARN `ApplicationMaster`
Spark RAPIDS	You can run RAPIDS with the configured Java version for Spark.
Iceberg	You can run Iceberg with the configured Java version of the application that is using it.
Delta	You can run Delta with the configured Java version of the application that is using it.
Hudi	You can run Hudi with the configured Java version of the application that is using it.
Hadoop	To update the JVM for Hadoop, modify `hadoop-env` . For examples, see [Override the JVM](#).
Hive	To set the Java version to 11 or 17 for Hive, configure the Hadoop JVM setting to the Java version that you want to use.

Application	Java configuration notes
HBase	To update the JVM for HBase, modify hbase-env . By default, Amazon EMR sets the HBase JVM based on the JVM configuration for Hadoop unless you override the settings in hbase-env . For examples, see Override the JVM.
Flink	To update the JVM for Flink, modify flink-conf . By default, Amazon EMR sets the Flink JVM based on the JVM configuration for Hadoop unless you override the settings in flink-conf . For more information, see Configure Flink to run with Java 11.
Oozie	To configure Oozie to run on Java 11 or 17, configure Oozie Server, the Oozie LauncherAM Launcher AM, and change your client-side executable and job configurations. You can also configure EmbeddedOozieServer to run on Java 17. For more information, see Configure Java version for Oozie.
Pig	Pig only supports Java 8. You can't use Java 11 or 17 with Hadoop and run Pig on the same cluster.

Override the JVM

To override the JVM setting for an Amazon EMR release - for example, to use Java 17 with a cluster that uses Amazon EMR release 6.12.0 - supply the `JAVA_HOME` setting to its environment classification, which is *`application`*`-env` for all applications except Flink. For Flink, the environment classification is `flink-conf`. For steps to configure the Java runtime with Flink, see Configure Flink to run with Java 11.

Topics

- [Override the JVM setting with Apache Spark]
- [Override the JVM setting with Apache HBase]
- [Override the JVM setting with Apache Hadoop and Hive]

Override the JVM setting with Apache Spark

When you use Spark with Amazon EMR releases 6.12 and higher, if you write a driver for submission in cluster mode, the driver uses Java 8, but you can set the environment so that the executors use Java 11 or 17. And when you use Spark with Amazon EMR releases lower than 5.x, and you write a driver for submission in cluster mode, the driver uses Java 7. However, you can set the environment to ensure that the executors use Java 8.

To override the JVM for Spark, we recommend that you set both the Hadoop and Spark classifications.

```
{
"Classification": "hadoop-env",
        "Configurations": [
            {
"Classification": "export",
                "Configurations": [],
                "Properties": {
"JAVA_HOME": "/usr/lib/jvm/java-1.8.0"
                }
            }
        ],
        "Properties": {}
    },
    {
"Classification": "spark-env",
        "Configurations": [
            {
"Classification": "export",
                "Configurations": [],
                "Properties": {
"JAVA_HOME": "/usr/lib/jvm/java-1.8.0"
                }
            }
        ],
        "Properties": {}
```

```
    }
```

Override the JVM setting with Apache HBase

To configure HBase to use Java 11, you can set the following configuration when you launch the cluster.

```
[
  {
    "Classification": "hbase-env",
    "Configurations": [
      {
        "Classification": "export",
        "Configurations": [],
        "Properties": {
          "JAVA_HOME": "/usr/lib/jvm/jre-11"
        }
      }
    ],
    "Properties": {}
  }
]
```

Override the JVM setting with Apache Hadoop and Hive

The following example shows how to set the JVM to version 17 for Hadoop and Hive.

```
[
    {
        "Classification": "hadoop-env",
            "Configurations": [
                {
                    "Classification": "export",
                    "Configurations": [],
                    "Properties": {
                        "JAVA_HOME": "/usr/lib/jvm/jre-17"
                    }
                }
        ],
        "Properties": {}
    }
```

```
]
```

Service ports

The following are YARN and HDFS service ports. These settings reflect Hadoop defaults. Other application services are hosted at default ports unless otherwise documented. For more information, see the application's project documentation.

Port settings for YARN and HDFS

Setting	Hostname/Port
`fs.default.name`	default (`hdfs://`*`emrDeterminedIP`*`:8020`)
`dfs.datanode.address`	default (`0.0.0.0:50010`)
`dfs.datanode.http.address`	default (`0.0.0.0:50075`)
`dfs.datanode.https.address`	default (`0.0.0.0:50475`)
`dfs.datanode.ipc.address`	default (`0.0.0.0:50020`)
`dfs.http.address`	default (`0.0.0.0:50070`)
`dfs.https.address`	default (`0.0.0.0:50470`)
`dfs.secondary.http.address`	default (`0.0.0.0:50090`)
`yarn.nodemanager.address`	default (`${yarn.nodemanager.hostname}:0`)
`yarn.nodemanager.localizer.address`	default (`${yarn.nodemanager.hostname}:8040`)
`yarn.nodemanager.webapp.address`	default (`${yarn.nodemanager.hostname}:8042`)
`yarn.resourcemanager.address`	default (`${yarn.resourcemanager.hostname}:8032`)

Setting	Hostname/Port
yarn.resourcemanager.admin.address	default (${yarn.resourcemanager.hostname}:8033)
yarn.resourcemanager.resource-tracker.address	default (${yarn.resourcemanager.hostname}:8031)
yarn.resourcemanager.scheduler.address	default (${yarn.resourcemanager.hostname}:8030)
yarn.resourcemanager.webapp.address	default (${yarn.resourcemanager.hostname}:8088)
yarn.web-proxy.address	default (no-value)
yarn.resourcemanager.hostname	*emrDeterminedIP*

Note

The term *emrDeterminedIP* is an IP address that is generated by the Amazon EMR control plane. In the newer version, this convention has been removed, except for the yarn.resourcemanager.hostname and fs.default.name settings.

Application users

Applications run processes as their own user. For example, Hive JVMs run as user hive, MapReduce JVMs run as mapred, and so on. This is demonstrated in the following process status example.

```
USER       PID %CPU %MEM    VSZ   RSS TTY      STAT START   TIME COMMAND
hive      6452  0.2  0.7 853684 218520 ?       Sl   16:32   0:13 /usr/lib/jvm/
java-openjdk/bin/java -Xmx256m -Dhive.log.dir=/var/log/hive -Dhive.log.file=hive-
metastore.log -Dhive.log.threshold=INFO -Dhadoop.log.dir=/usr/lib/hadoop
hive      6557  0.2  0.6 849508 202396 ?       Sl   16:32   0:09 /usr/lib/jvm/java-
openjdk/bin/java -Xmx256m -Dhive.log.dir=/var/log/hive -Dhive.log.file=hive-server2.log
 -Dhive.log.threshold=INFO -Dhadoop.log.dir=/usr/lib/hadoop/l
```

```
hbase     6716  0.1  1.0 1755516 336600 ?      Sl   Jun21   2:20 /usr/lib/jvm/java-
openjdk/bin/java -Dproc_master -XX:OnOutOfMemoryError=kill -9 %p -Xmx1024m -ea -XX:
+UseConcMarkSweepGC -XX:+CMSIncrementalMode -Dhbase.log.dir=/var/
hbase     6871  0.0  0.7 1672196 237648 ?      Sl   Jun21   0:46 /usr/lib/jvm/java-
openjdk/bin/java -Dproc_thrift -XX:OnOutOfMemoryError=kill -9 %p -Xmx1024m -ea -XX:
+UseConcMarkSweepGC -XX:+CMSIncrementalMode -Dhbase.log.dir=/var/
hdfs      7491  0.4  1.0 1719476 309820 ?      Sl   16:32   0:22 /usr/lib/jvm/java-
openjdk/bin/java -Dproc_namenode -Xmx1000m -Dhadoop.log.dir=/var/log/hadoop-hdfs -
Dhadoop.log.file=hadoop-hdfs-namenode-ip-10-71-203-213.log -Dhadoo
yarn      8524  0.1  0.6 1626164 211300 ?      Sl   16:33   0:05 /usr/lib/jvm/java-
openjdk/bin/java -Dproc_proxyserver -Xmx1000m -Dhadoop.log.dir=/var/log/hadoop-yarn -
Dyarn.log.dir=/var/log/hadoop-yarn -Dhadoop.log.file=yarn-yarn-
yarn      8646  1.0  1.2 1876916 385308 ?      Sl   16:33   0:46 /usr/lib/jvm/java-
openjdk/bin/java -Dproc_resourcemanager -Xmx1000m -Dhadoop.log.dir=/var/log/hadoop-yarn
 -Dyarn.log.dir=/var/log/hadoop-yarn -Dhadoop.log.file=yarn-y
mapred    9265  0.2  0.8 1666628 260484 ?      Sl   16:33   0:12 /usr/lib/jvm/java-
openjdk/bin/java -Dproc_historyserver -Xmx1000m -Dhadoop.log.dir=/usr/lib/hadoop/logs -
Dhadoop.log.file=hadoop.log -Dhadoop.home.dir=/usr/lib/hadoop
```

Checking dependencies using the Amazon EMR artifact repository

You can use the Amazon EMR artifact repository to build Apache Hive and Apache Hadoop job code against the exact versions of libraries and dependencies that are available with specific Amazon EMR release versions, beginning with Amazon EMR release version 5.18.0. Building against Amazon EMR artifacts in the repository helps avoid runtime class path issues by ensuring that the versions of the libraries that the job is built against are exactly the same versions provided at runtime on the cluster. Currently, Amazon EMR artifacts are only available for Maven builds.

To access the artifact repository, add the repository URL to your Maven settings file or to a specific project's `pom.xml` configuration file. You can then specify the dependencies in your project configuration. For dependency versions, use the version listed under *Component Versions* for the desired release on Amazon EMR 5.x release versions. For example, component versions for the most recent Amazon EMR release are available at the section called "5.36.2 component versions". If an artifact for your project is not listed under *Component Versions*, specify the version that is listed for Hive and Hadoop in that release. For example, for Hadoop components in Amazon EMR release version 5.18.0, the version is `2.8.4-amzn-1`.

The artifact repository URL has the following syntax:

```
https://s3-endpoint/region-ID-emr-artifacts/emr-release-label/repos/maven/
```

- *`s3-endpoint`* is the Amazon Simple Storage Service (Amazon S3) endpoint of the region for the repository and *`region-ID`* is the corresponding region. For example, `s3.us-west-1.amazonaws.com` and `us-west-1`. For more information, see Amazon S3 endpoints in the *Amazon Web Services General Reference*. There is no difference in artifacts between regions, so you can specify the most convenient region for your development environment.
- *`emr-release-label`* is the release label for the Amazon EMR cluster that will run your code. Release labels are in the form `emr-x.x.x`, such as, `emr-5.36.2`. An EMR release series may include multiple releases. For example, if you're using EMR release version 5.24.1, use the first EMR release label within the 5.24 series, `emr-5.24.0`, in the artifact repository URL:

```
https://s3-endpoint/region-ID-emr-artifacts/emr-5.24.0/repos/maven/
```

Example Configuration for Maven pom.xml

The pom.xml example below configures a Maven project to build against the emr-5.18.0 Apache Hadoop and Apache Hive artifacts, using the artifact repository in us-west-1. Snapshot versions are not available in the artifact repository, so snapshots are disabled in the pom.xml. Ellipses (`...`) in the example below indicate omission of other configuration parameters. Do not copy these into your Maven project.

```
<project>
 ...
 <repositories>
  ...
  <repository>
   <id>emr-5.18.0-artifacts</id>
   <name>EMR 5.18.0 Releases Repository</name>
   <releases>
    <enabled>true</enabled>
   </releases>
   <snapshots>
    <enabled>false</enabled>
   </snapshots>
   <url>https://s3.us-west-1.amazonaws.com/us-west-1-emr-artifacts/emr-5.18.0/repos/
maven/</url>
  </repository>
  ...
 </repositories>
 ...
 <dependencies>
  ...
  <dependency>
   <groupId>org.apache.hive</groupId>
   <artifactId>hive-exec</artifactId>
   <version>2.3.3-amzn-2</version>
  </dependency>
  <dependency>
   <groupId>org.apache.hadoop</groupId>
   <artifactId>hadoop-common</artifactId>
   <version>2.8.4-amzn-1</version>
  </dependency>
  ...
 </dependencies>
```

```
</project>
```

EMR File System (EMRFS)

The EMR File System (EMRFS) is an implementation of HDFS that all Amazon EMR clusters use for reading and writing regular files from Amazon EMR directly to Amazon S3. EMRFS provides the convenience of storing persistent data in Amazon S3 for use with Hadoop while also providing features like data encryption.

Data encryption allows you to encrypt objects that EMRFS writes to Amazon S3, and enables EMRFS to work with encrypted objects in Amazon S3. If you're using Amazon EMR release version 4.8.0 or later, you can use security configurations to set up encryption for EMRFS objects in Amazon S3, along with other encryption settings. For more information, see Encryption options. If you use an earlier release version of Amazon EMR, you can manually configure encryption settings. For more information, see Specifying Amazon S3 encryption using EMRFS properties.

Amazon S3 offers strong read-after write consistency for all GET, PUT, and LIST operations across all AWS Regions. This means that what you write using EMRFS is what you'll read from Amazon S3, with no impact on performance. For more information, see Amazon S3 data consistency model.

When using Amazon EMR release version 5.10.0 or later, you can use different IAM roles for EMRFS requests to Amazon S3 based on cluster users, groups, or the location of EMRFS data in Amazon S3. For more information, see Configure IAM roles for EMRFS requests to Amazon S3.

Warning

Before turning on speculative execution for Amazon EMR clusters running Apache Spark jobs, please review the following information.

EMRFS includes the EMRFS S3-optimized committer, an OutputCommitter implementation that is optimized for writing files to Amazon S3 when using EMRFS. If you turn on the Apache Spark speculative execution feature with applications that write data to Amazon S3 and do not use the EMRFS S3-optimized committer, you may encounter data correctness issues described in SPARK-10063. This can occur if you are using Amazon EMR versions earlier than Amazon EMR release 5.19, or if you are writing files to Amazon S3 with formats such as ORC and CSV. These formats aren't supported by the EMRFS S3-optimized committer. For a complete list of requirements for using the EMRFS S3-optimized committer, see Requirements for the EMRFS S3-optimized committer.

EMRFS direct write is typically used when the EMRFS S3-optimized committer is not supported, such as when writing the following:

- An output format other than Parquet, such as ORC or text.
- Hadoop files using the Spark RDD API.
- Parquet using Hive SerDe. See Hive metastore Parquet table conversion.

EMRFS direct write is not used in the following scenarios:

- When the EMRFS S3-optimized committer is enabled. See Requirements for the EMRFS S3-optimized committer.
- When writing dynamic partitions with partitionOverwriteMode set to dynamic.
- When writing to custom partition locations, such as locations that do not conform to the Hive default partition location convention.
- When using file systems other than EMRFS, such as writing to HDFS or using the S3A file system.

To determine whether your application uses direct write in Amazon EMR 5.14.0 or later, enable Spark INFO logging. If a log line containing the text "Direct Write: ENABLED" is present in either Spark driver logs or Spark executor container logs, then your Spark application wrote using direct write.
By default, speculative execution is turned OFF on Amazon EMRclusters. We highly recommend that you do not turn speculative execution on if both of these conditons are true:

- You are writing data to Amazon S3.
- Data is written in a format other than Apache Parquet or in Apache Parquet format not using the EMRFS S3-optimized committer.

If you turn on Spark speculative execution and write data to Amazon S3 using EMRFS direct write, you may experience intermittent data loss. When you write data to HDFS, or write data in Parquet using the EMRFS S3-optimized committer, Amazon EMR does not use direct write and this issue does not occur.
If you need to write data in formats that use EMRFS direct write from Spark to Amazon S3 and use speculative execution, we recommend writing to HDFS and then transferring output files to Amazon S3 using S3DistCP.

Topics

Consistent view

> **Warning**
>
> On June 1, 2023, EMRFS consistent view will reach end of standard support for future Amazon EMR releases. EMRFS consistent view will continue to work for existing releases.

With the release of Amazon S3 strong read-after-write consistency on December 1, 2020, you no longer need to use EMRFS consistent view (EMRFS CV) with your Amazon EMR clusters. EMRFS CV is an optional feature that allows Amazon EMR clusters to check for list and read-after-write consistency for Amazon S3 objects. When you create a cluster and EMRFS CV is turned on, Amazon EMR creates an Amazon DynamoDB database to store object metadata that it uses to track list and read-after-write consistency for S3 objects. You can now turn off EMRFS CV and delete the DynamoDB database that it uses so that you don't accrue additional costs. The following procedures explain how to check for the CV feature, turn it off, and delete the DynamoDB database that the feature uses.

To check if you're using the EMRFS CV feature

1. Navigate to the **Configuration** tab. If your cluster has the following configuration, it uses EMRFS CV.

   ```
   Classification=emrfs-site,Property=fs.s3.consistent,Value=true
   ```

2. Alternatively, use the AWS CLI to describe your cluster with the `describe-cluster` API. If the output contains `fs.s3.consistent: true`, your cluster uses EMRFS CV.

To turn off EMRFS CV on your Amazon EMR clusters

To turn off the EMRFS CV feature, use one of the following three options. You should test these options in your testing environment before applying them to your production environments.

1. **To stop your existing cluster and start a new cluster without EMRFS CV options.**

    a. Before you stop your cluster, ensure that you back up your data and notify your users.

    b. To stop your cluster, follow the instructions in Terminate a cluster.

    c. If you use the Amazon EMR console to create new cluster, navigate to **Advanced Options**. In the **Edit software settings** section, deselect the option to turn on EMRFS CV. If the check box for **EMRFS consistent view** is available, keep it unchecked.

    d. If you use AWS CLI to create a new cluster with the `create-cluster` API, don't use the `--emrfs` option, which turns on EMRFS CV.

    e. If you use an SDK or AWS CloudFormation to create a new cluster, don't use any of the configurations listed in Configure consistent view.

2. **To clone a cluster and remove EMRFS CV**

    a. In the Amazon EMR console, choose the cluster that uses EMRFS CV.

    b. At the top of the **Cluster Details** page, choose **Clone**.

    c. Choose **Previous** and navigate to **Step 1: Software and Steps**.

    d. In **Edit software settings**, remove EMRFS CV. In **Edit configuration**, delete the following configurations in the `emrfs-site` classification. If you're loading JSON from a S3 bucket, you must modify your S3 object.

```
[
 {"classification":
  "emrfs-site",
  "properties": {
   "fs.s3.consistent.retryPeriodSeconds":"10",
   "fs.s3.consistent":"true",
   "fs.s3.consistent.retryCount":"5",
   "fs.s3.consistent.metadata.tableName":"EmrFSMetadata"
  }
 }
]
```

3. **To remove EMRFS CV from a cluster that uses instance groups**

   a. Use the following command to check if a single EMR cluster uses the DynamoDB table that is associated with EMRFS CV, or if multiple clusters share the table. The table name is specified in `fs.s3.consistent.metadata.tableName`, as described in [Configure consistent view](). The default table name used by EMRFS CV is `EmrFSMetadata`.

```
aws emr describe-cluster --cluster-id j-XXXXX | grep
 fs.s3.consistent.metadata.tableName
```

   b. If your cluster doesn't share your DynamoDB database with another cluster, use the following command to reconfigure the cluster and deactivate EMRFS CV. For more information, see [Reconfigure an instance group in a running cluster]().

```
aws emr modify-instance-groups --cli-input-json file://disable-emrfs-1.json
```

   This command opens the file you want to modify. Modify the file with the following configurations.

```
{
 "ClusterId": "j-xxxx",
 "InstanceGroups": [
  {
   "InstanceGroupId": "ig-xxxx",
   "Configurations": [
    {
     "Classification": "emrfs-site",
     "Properties": {
      "fs.s3.consistent": "false"
     },
     "Configurations": []
    }
   ]
  }
 ]
}
```

   c. If your cluster shares the DynamoDB table with another cluster, turn off EMRFS CV on all clusters at a time when no clusters modify any objects in the shared S3 location.

To delete Amazon DynamoDB resources associated with EMRFS CV

After you remove EMRFS CV from your Amazon EMR clusters, delete the DynamoDB resources associated with EMRFS CV. Until you do so, you continue to incur DynamoDB charges associated with EMRFS CV.

1. Check the CloudWatch metrics for your DynamoDB table and confirm that the table isn't used by any clusters.
2. Delete the DynamoDB table.

```
aws dynamodb delete-table --table-name <your-table-name>
```

To delete Amazon SQS resources associated with EMRFS CV

1. If you configured your cluster to push inconsistency notifications to Amazon SQS, you can delete all SQS queues.
2. Find the Amazon SQS queue name specified in `fs.s3.consistent.notification.SQS.queueName`, as described in [Configure consistent view](). The default queue name format is `EMRFS-Inconsistency-`*`<j-cluster ID>`*.

```
aws sqs list-queues | grep 'EMRFS-Inconsistency'
aws sqs delete-queue -queue-url <your-queue-url>
```

To stop using the EMRFS CLI

- The [EMRFS CLI]() manages the metadata that EMRFS CV generates. As standard support for EMRFS CV reaches its end in future releases of Amazon EMR, support for the EMRFS CLI will also reach its end.

Topics

- Configure consistent view
- EMRFS CLI Command Reference

Enable consistent view

You can enable Amazon S3 server-side encryption or consistent view for EMRFS using the AWS Management Console, AWS CLI, or the `emrfs-site` configuration classification.

To configure consistent view using the console

1. Navigate to the new Amazon EMR console and select **Switch to the old console** from the side navigation. For more information on what to expect when you switch to the old console, see Using the old console.
2. Choose **Create cluster**, **Go to advanced options**.
3. Choose settings for **Step 1: Software and Steps** and **Step 2: Hardware**.
4. For **Step 3: General Cluster Settings**, under **Additional Options**, choose **EMRFS consistent view**.
5. For **EMRFS Metadata store**, type the name of your metadata store. The default value is **`EmrFSMetadata`**. If the EmrFSMetadata table does not exist, it is created for you in DynamoDB.

   > **Note**
   >
   > Amazon EMR does not automatically remove the EMRFS metadata from DynamoDB when the cluster is terminated.

6. For **Number of retries**, type an integer value. If an inconsistency is detected, EMRFS tries to call Amazon S3 this number of times. The default value is **5**.
7. For **Retry period (in seconds)**, type an integer value. This is the amount of time that EMRFS waits between retry attempts. The default value is **`10`**.

   > **Note**
   >
   > Subsequent retries use an exponential backoff.

To launch a cluster with consistent view enabled using the AWS CLI

We recommend that you install the current version of AWS CLI. To download the latest release, see https://aws.amazon.com/cli/.

- **Note**

  Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --instance-type m5.xlarge --instance-count 3 --emrfs
 Consistent=true \
--release-label emr-7.2.0 --ec2-attributes KeyName=myKey
```

To check if consistent view is enabled using the AWS Management Console

- To check whether consistent view is enabled in the console, navigate to the **Cluster List** and select your cluster name to view **Cluster Details**. The "EMRFS consistent view" field has a value of `Enabled` or `Disabled`.

To check if consistent view is enabled by examining the `emrfs-site.xml` file

- You can check if consistency is enabled by inspecting the `emrfs-site.xml` configuration file on the master node of the cluster. If the Boolean value for `fs.s3.consistent` is set to `true` then consistent view is enabled for file system operations involving Amazon S3.

Understanding how EMRFS consistent view tracks objects in Amazon S3

EMRFS creates a consistent view of objects in Amazon S3 by adding information about those objects to the EMRFS metadata. EMRFS adds these listings to its metadata when:

- An object written by EMRFS during the course of an Amazon EMR job.
- An object is synced with or imported to EMRFS metadata by using the EMRFS CLI.

Objects read by EMRFS are not automatically added to the metadata. When EMRFS deletes an object, a listing still remains in the metadata with a deleted state until that listing is purged using the EMRFS CLI. To learn more about the CLI, see EMRFS CLI Command Reference. For more information about purging listings in the EMRFS metadata, see EMRFS consistent view metadata.

For every Amazon S3 operation, EMRFS checks the metadata for information about the set of objects in consistent view. If EMRFS finds that Amazon S3 is inconsistent during one of these operations, it retries the operation according to parameters defined in `emrfs-site` configuration properties. After EMRFS exhausts the retries, it either throws a `ConsistencyException` or logs the exception and continue the workflow. For more information about retry logic, see Retry logic. You can find `ConsistencyExceptions` in your logs, for example:

- listStatus: No Amazon S3 object for metadata item `/S3_bucket/dir/object`
- getFileStatus: Key `dir/file` is present in metadata but not Amazon S3

If you delete an object directly from Amazon S3 that EMRFS consistent view tracks, EMRFS treats that object as inconsistent because it is still listed in the metadata as present in Amazon S3. If your metadata becomes out of sync with the objects EMRFS tracks in Amazon S3, you can use the **sync** sub-command of the EMRFS CLI to reset metadata so that it reflects Amazon S3. To discover discrepancies between metadata and Amazon S3, use the **diff**. Finally, EMRFS only has a consistent view of the objects referenced in the metadata; there can be other objects in the same Amazon S3 path that are not being tracked. When EMRFS lists the objects in an Amazon S3 path, it returns the superset of the objects being tracked in the metadata and those in that Amazon S3 path.

Retry logic

EMRFS tries to verify list consistency for objects tracked in its metadata for a specific number of retries. The default is 5. In the case where the number of retries is exceeded the originating job returns a failure unless `fs.s3.consistent.throwExceptionOnInconsistency` is set to `false`, where it will only log the objects tracked as inconsistent. EMRFS uses an exponential backoff retry policy by default but you can also set it to a fixed policy. Users may also want to retry for a certain period of time before proceeding with the rest of their job without throwing an exception. They can achieve this by setting `fs.s3.consistent.throwExceptionOnInconsistency` to `false`, `fs.s3.consistent.retryPolicyType` to `fixed`, and `fs.s3.consistent.retryPeriodSeconds` for the desired value. The following example creates a cluster with consistency enabled, which logs inconsistencies and sets a fixed retry interval of 10 seconds:

Example Setting retry period to a fixed amount

```
aws emr create-cluster --release-label emr-7.2.0 \
--instance-type m5.xlarge --instance-count 1 \
--emrfs Consistent=true,Args=[fs.s3.consistent.throwExceptionOnInconsistency=false,
 fs.s3.consistent.retryPolicyType=fixed,fs.s3.consistent.retryPeriodSeconds=10] --ec2-
attributes KeyName=myKey
```

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

For more information, see Consistent view.

EMRFS configurations for IMDS get region calls

EMRFS relies on the IMDS (instance metadata service) to get instance region and Amazon S3, DynamoDB, or AWS KMS endpoints. However, IMDS has a limit on how many requests it can handle, and requests that exceed that limit will fail. This IMDS limit can cause EMRFS failures to initialize and cause the query or command to fail. You can use the following randomized exponential backoff retry mechanism and a fallback region configuration properties in emrfs-site.xml to address the scenario where all retries fail.

```
<property>
    <name>fs.s3.region.retryCount</name>
    <value>3</value>
    <description>
    Maximum retries that would be attempted to get AWS region.
    </description>
</property>
<property>
    <name>fs.s3.region.retryPeriodSeconds</name>
    <value>3</value>
    <description>
    Base sleep time in second for each get-region retry.
```

```
    </description>
</property>
<property>
    <name>fs.s3.region.fallback</name>
    <value>us-east-1</value>
    <description>
    Fallback to this region after maximum retries for getting AWS region have been
 reached.
    </description>
</property>
```

EMRFS consistent view metadata

EMRFS consistent view tracks consistency using a DynamoDB table to track objects in Amazon S3 that have been synced with or created by EMRFS. The metadata is used to track all operations (read, write, update, and copy), and no actual content is stored in it. This metadata is used to validate whether the objects or metadata received from Amazon S3 matches what is expected. This confirmation gives EMRFS the ability to check list consistency and read-after-write consistency for new objects EMRFS writes to Amazon S3 or objects synced with EMRFS. Multiple clusters can share the same metadata.

How to add entries to metadata

You can use the `sync` or `import` subcommands to add entries to metadata. `sync` reflects the state of the Amazon S3 objects in a path, while `import` is used strictly to add new entries to the metadata. For more information, see EMRFS CLI Command Reference.

How to check differences between metadata and objects in Amazon S3

To check for differences between the metadata and Amazon S3, use the `diff` subcommand of the EMRFS CLI. For more information, see EMRFS CLI Command Reference.

How to know if metadata operations are being throttled

EMRFS sets default throughput capacity limits on the metadata for its read and write operations at 500 and 100 units, respectively. Large numbers of objects or buckets may cause operations to exceed this capacity, at which point DynamoDB will throttle operations. For example, an application may cause EMRFS to throw a `ProvisionedThroughputExceededException` if you perform an operation that exceeds these capacity limits. Upon throttling, the EMRFS CLI tool attempts to retry writing to the DynamoDB table using exponential backoff until the operation finishes or when it reaches the maximum retry value for writing objects from Amazon EMR to Amazon S3.

You can configure your own throughput capacity limits. However, DynamoDB has strict partition limits of 3000 read capacity units (RCUs) and 1000 write capacity units (WCUs) per second for read and write operations. To avoid `sync` failures caused by throttling, we recommend you limit throughput for read operations to fewer than 3000 RCUs and write operations to fewer than 1000 WCUs. For instructions on setting custom throughput capacity limits, see Configure consistent view.

You can also view Amazon CloudWatch metrics for your EMRFS metadata in the DynamoDB console where you can see the number of throttled read and write requests. If you do have a non-zero value for throttled requests, your application may potentially benefit from increasing allocated throughput capacity for read or write operations. You may also realize a performance benefit if you see that your operations are approaching the maximum allocated throughput capacity in reads or writes for an extended period of time.

Throughput characteristics for notable EMRFS operations

The default for read and write operations is 400 and 100 throughput capacity units, respectively. The following performance characteristics give you an idea of what throughput is required for certain operations. These tests were performed using a single-node `m3.large` cluster. All operations were single threaded. Performance differs greatly based on particular application characteristics and it may take experimentation to optimize file system operations.

Operation	Average read-per-second	Average write-per-second
create (object)	26.79	6.70
delete (object)	10.79	10.79
delete (directory containing 1000 objects)	21.79	338.40
getFileStatus (object)	34.70	0
getFileStatus (directory)	19.96	0

Operation	Average read-per-second	Average write-per-second
listStatus (directory containing 1 object)	43.31	0
listStatus (directory containing 10 objects)	44.34	0
listStatus (directory containing 100 objects)	84.44	0
listStatus (directory containing 1,000 objects)	308.81	0
listStatus (directory containing 10,000 objects)	416.05	0
listStatus (directory containing 100,000 objects)	823.56	0
listStatus (directory containing 1M objects)	882.36	0
mkdir (continuous for 120 seconds)	24.18	4.03
mkdir	12.59	0
rename (object)	19.53	4.88
rename (directory containing 1000 objects)	23.22	339.34

To submit a step that purges old data from your metadata store

Users may wish to remove particular entries in the DynamoDB-based metadata. This can help reduce storage costs associated with the table. Users have the ability to manually or programmatically purge particular entries by using the EMRFS CLI `delete` subcommand. However, if you delete entries from the metadata, EMRFS no longer makes any checks for consistency.

Programmatically purging after the completion of a job can be done by submitting a final step to your cluster, which executes a command on the EMRFS CLI. For instance, type the following command to submit a step to your cluster to delete all entries older than two days.

```
aws emr add-steps --cluster-id j-2AL4XXXXXX5T9 --steps Name="emrfsCLI",Jar="command-runner.jar",Args=["emrfs","delete","--time","2","--time-unit","days"]
{
    "StepIds": [
        "s-B12345678902"
    ]
}
```

Use the StepId value returned to check the logs for the result of the operation.

Configure consistency notifications for CloudWatch and Amazon SQS

You can enable CloudWatch metrics and Amazon SQS messages in EMRFS for Amazon S3 eventual consistency issues.

CloudWatch

When CloudWatch metrics are enabled, a metric named **Inconsistency** is pushed each time a `FileSystem` API call fails due to Amazon S3 eventual consistency.

To view CloudWatch metrics for Amazon S3 eventual consistency issues

To view the **Inconsistency** metric in the CloudWatch console, select the EMRFS metrics and then select a **JobFlowId/Metric Name** pair. For example: `j-162XXXXXXM2CU ListStatus`, `j-162XXXXXXM2CU GetFileStatus`, and so on.

1. Open the CloudWatch console at https://console.aws.amazon.com/cloudwatch/.
2. In the **Dashboard**, in the **Metrics** section, choose **EMRFS**.
3. In the **Job Flow Metrics** pane, select one or more **JobFlowId/Metric Name** pairs. A graphical representation of the metrics appears in the window below.

Amazon SQS

When Amazon SQS notifications are enabled, an Amazon SQS queue with the name `EMRFS-Inconsistency-<jobFlowId>` is created when EMRFS is initialized. Amazon SQS messages are pushed into the queue when a `FileSystem` API call fails due to Amazon S3 eventual consistency. The message contains information such as JobFlowId, API, a list of inconsistent paths, a stack trace, and so on. Messages can be read using the Amazon SQS console or using the EMRFS `read-sqs` command.

To manage Amazon SQS messages for Amazon S3 eventual consistency issues

Amazon SQS messages for Amazon S3 eventual consistency issues can be read using the EMRFS CLI. To read messages from an EMRFS Amazon SQS queue, type the `read-sqs` command and specify an output location on the master node's local file system for the resulting output file.

You can also delete an EMRFS Amazon SQS queue using the `delete-sqs` command.

1. To read messages from an Amazon SQS queue, type the following command. Replace *`queuename`* with the name of the Amazon SQS queue that you configured and replace *`/path/filename`* with the path to the output file:

   ```
   emrfs read-sqs --queue-name queuename --output-file /path/filename
   ```

   For example, to read and output Amazon SQS messages from the default queue, type:

   ```
   emrfs read-sqs --queue-name EMRFS-Inconsistency-j-162XXXXXXM2CU --output-file /
   path/filename
   ```

   Note

   You can also use the `-q` and `-o` shortcuts instead of `--queue-name` and `--output-file` respectively.

2. To delete an Amazon SQS queue, type the following command:

   ```
   emrfs delete-sqs --queue-name queuename
   ```

   For example, to delete the default queue, type:

   ```
   emrfs delete-sqs --queue-name EMRFS-Inconsistency-j-162XXXXXXM2CU
   ```

> **Note**
>
> You can also use the `-q` shortcut instead of `--queue-name`.

Configure consistent view

You can configure additional settings for consistent view by providing them using configuration properties for `emrfs-site` properties. For example, you can choose a different default DynamoDB throughput by supplying the following arguments to the CLI `--emrfs` option, using the emrfs-site configuration classification (Amazon EMR release version 4.x and later only), or a bootstrap action to configure the emrfs-site.xml file on the master node:

Example Changing default metadata read and write values at cluster launch

```
aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge \
--emrfs Consistent=true,Args=[fs.s3.consistent.metadata.read.capacity=600,\
fs.s3.consistent.metadata.write.capacity=300] --ec2-attributes KeyName=myKey
```

Alternatively, use the following configuration file and save it locally or in Amazon S3:

```
[
    {
      "Classification": "emrfs-site",
      "Properties": {
        "fs.s3.consistent.metadata.read.capacity": "600",
        "fs.s3.consistent.metadata.write.capacity": "300"
      }
    }
 ]
```

Use the configuration you created with the following syntax:

```
aws emr create-cluster --release-label emr-7.2.0 --applications Name=Hive \
--instance-type m5.xlarge --instance-count 2 --configurations file://./myConfig.json
```

Note

Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

The following options can be set using configurations or AWS CLI `--emrfs` arguments. For information about those arguments, see the AWS CLI Command Reference.

`emrfs-site.xml` Properties for consistent view

Property	Default value	Description
`fs.s3.consistent`	**`false`**	When set to **`true`**, this property configures EMRFS to use DynamoDB to provide consistency.
`fs.s3.consistent.retryPolicyType`	**`exponential`**	This property identifies the policy to use when retrying for consistency issues. Options include: exponential, fixed, or none.
`fs.s3.consistent.retryPeriodSeconds`	**`1`**	This property sets the length of time to wait between consistency retry attempts.
`fs.s3.consistent.retryCount`	**`10`**	This property sets the maximum number of retries when inconsistency is detected.
`fs.s3.consistent.throwExceptionOnInconsistency`	**`true`**	This property determines whether to throw or log a consistency exception. When set to **`true`**, a `ConsistencyException` is thrown.

Property	Default value	Description
fs.s3.consistent.metadata.autoCreate	**true**	When set to **true**, this property enables automatic creation of metadata tables.
fs.s3.consistent.metadata.etag.verification.enabled	**true**	With Amazon EMR 5.29.0, this property is enabled by default. When enabled, EMRFS uses S3 ETags to verify that objects being read are the latest available version. This feature is helpful for read-after-update use cases in which files on S3 are being overwritten while retaining the same name. This ETag verification capability currently does not work with S3 Select.
fs.s3.consistent.metadata.tableName	**EmrFSMetadata**	This property specifies the name of the metadata table in DynamoDB.
fs.s3.consistent.metadata.read.capacity	**500**	This property specifies the DynamoDB read capacity to provision when the metadata table is created.
fs.s3.consistent.metadata.write.capacity	**100**	This property specifies the DynamoDB write capacity to provision when the metadata table is created.

Property	Default value	Description
fs.s3.consistent.fastList	**true**	When set to **true**, this property uses multiple threads to list a directory (when necessary). Consistency must be enabled in order to use this property.
fs.s3.consistent.fastList.prefetchMetadata	**false**	When set to **true**, this property enables metadata prefetching for directories containing more than 20,000 items.
fs.s3.consistent.notification.CloudWatch	**false**	When set to **true**, CloudWatch metrics are enabled for FileSystem API calls that fail due to Amazon S3 eventual consistency issues.
fs.s3.consistent.notification.SQS	**false**	When set to **true**, eventual consistency notifications are pushed to an Amazon SQS queue.
fs.s3.consistent.notification.SQS.queueName	**EMRFS-Inconsistency-<jobFlowId>**	Changing this property allows you to specify your own SQS queue name for messages regarding Amazon S3 eventual consistency issues.

Property	Default value	Description
`fs.s3.consistent.notification.SQS.customMsg`	**none**	This property allows you to specify custom information included in SQS messages regarding Amazon S3 eventual consistency issues. If a value is not specified for this property, the corresponding field in the message is empty.
`fs.s3.consistent.dynamodb.endpoint`	**none**	This property allows you to specify a custom DynamoDB endpoint for your consistent view metadata.
`fs.s3.useRequesterPaysHeader`	**`false`**	When set to **`true`**, this property allows Amazon S3 requests to buckets with the request payer option enabled.

EMRFS CLI Command Reference

The EMRFS CLI is installed by default on all cluster master nodes created using Amazon EMR release version 3.2.1 or later. You can use the EMRFS CLI to manage the metadata for consistent view.

> **Note**
>
> The **emrfs** command is only supported with VT100 terminal emulation. However, it may work with other terminal emulator modes.

emrfs top-level command

The **emrfs** top-level command supports the following structure.

```
emrfs [describe-metadata | set-metadata-capacity | delete-metadata | create-metadata |
 \
list-metadata-stores | diff | delete | sync | import ] [options] [arguments]
```

Specify [options], with or without [arguments] as described in the following table. For [options] specific to sub-commands (`describe-metadata`, `set-metadata-capacity`, etc.), see each sub-command below.

[Options] for emrfs

Option	Description	Required
`-a` *AWS_ACCESS_KEY_ID* | `--access-key` *AWS_ACCESS_KEY_ID*	The AWS access key you use to write objects to Amazon S3 and to create or access a metadata store in DynamoDB. By default, *AWS_ACCESS_KEY_ID* is set to the access key used to create the cluster.	No
`-s` *AWS_SECRET_ACCESS_KEY* | `--secret-key` *AWS_SECRET_ACCESS_KEY*	The AWS secret key associated with the access key you use to write objects to Amazon S3 and to create or access a metadata store in DynamoDB. By default, *AWS_SECRET_ACCESS_KEY* is set to the secret key associated with the access key used to create the cluster.	No
`-v` | `--verbose`	Makes output verbose.	No
`-h` | `--help`	Displays the help message for the `emrfs` command with a usage statement.	No

emrfs describe-metadata sub-command

[Options] for emrfs describe-metadata

Option	Description	Required

Option	Description	Required
-m *METADATA_NAME* | --metadata-name *METADATA_NAME*	*METADATA_NAME* is the name of the DynamoDB metadata table. If the *METADATA_NAME* argument is not supplied, the default value is `EmrFSMetadata`.	No

Example emrfs describe-metadata example

The following example describes the default metadata table.

```
$ emrfs describe-metadata
EmrFSMetadata
  read-capacity: 400
  write-capacity: 100
  status: ACTIVE
  approximate-item-count (6 hour delay): 12
```

emrfs set-metadata-capacity sub-command

[Options] for emrfs set-metadata-capacity

Option	Description	Required
-m *METADATA_NAME* | --metadata-name *METADATA_NAME*	*METADATA_NAME* is the name of the DynamoDB metadata table. If the *METADATA_NAME* argument is not supplied, the default value is `EmrFSMetadata`.	No
-r *READ_CAPACITY* | --read-capacity *READ_CAPACITY*	The requested read throughput capacity for the metadata table. If the *READ_CAPACITY* argument is not supplied, the default value is `400`.	No
-w *WRITE_CAPACITY* | --write-capacity *WRITE_CAPACITY*	The requested write throughput capacity for the metadata table. If the *WRITE_CAPACITY* argument is not supplied, the default value is `100`.	No

Example emrfs set-metadata-capacity example

The following example sets the read throughput capacity to 600 and the write capacity to 150 for a metadata table named EmrMetadataAlt.

```
$ emrfs set-metadata-capacity --metadata-name EmrMetadataAlt  --read-capacity 600 --
write-capacity 150
  read-capacity: 400
  write-capacity: 100
  status: UPDATING
  approximate-item-count (6 hour delay): 0
```

emrfs delete-metadata sub-command

[Options] for emrfs delete-metadata

Option	Description	Required
-m *METADATA_NAME* | --metadata-name *METADATA_NAME*	*METADATA_NAME* is the name of the DynamoDB metadata table. If the *METADATA_NAME* argument is not supplied, the default value is EmrFSMetadata .	No

Example emrfs delete-metadata example

The following example deletes the default metadata table.

```
$ emrfs delete-metadata
```

emrfs create-metadata sub-command

[Options] for emrfs create-metadata

Option	Description	Required
-m *METADATA_NAME* | --metadata-name *METADATA_NAME*	*METADATA_NAME* is the name of the DynamoDB metadata table. If the *METADATA_NAME* argument is not supplied, the default value is EmrFSMetadata .	No

Option	Description	Required
`-r READ_CAPACITY` | `--read-capacity READ_CAPACITY`	The requested read throughput capacity for the metadata table. If the *READ_CAPACITY* argument is not supplied, the default value is `400`.	No
`-w WRITE_CAPACITY` | `--write-capacity WRITE_CAPACITY`	The requested write throughput capacity for the metadata table. If the *WRITE_CAPACITY* argument is not supplied, the default value is `100`.	No

Example emrfs create-metadata example

The following example creates a metadata table named `EmrFSMetadataAlt`.

```
$ emrfs create-metadata -m EmrFSMetadataAlt
Creating metadata: EmrFSMetadataAlt
EmrFSMetadataAlt
  read-capacity: 400
  write-capacity: 100
  status: ACTIVE
  approximate-item-count (6 hour delay): 0
```

emrfs list-metadata-stores sub-command

The **emrfs list-metadata-stores** sub-command has no [options].

Example List-metadata-stores example

The following example lists your metadata tables.

```
$ emrfs list-metadata-stores
  EmrFSMetadata
```

emrfs diff sub-command

[Options] for emrfs diff

Option	Description	Required

Option	Description	Required
-m *METADATA_NAME* | --metadata-name *METADATA_NAME*	*METADATA_NAME* is the name of the DynamoDB metadata table. If the *METADATA_NAME* argument is not supplied, the default value is EmrFSMetadata .	No
s3://s3Path	The path to the Amazon S3 bucket to compare with the metadata table. Buckets sync recursively.	Yes

Example emrfs diff example

The following example compares the default metadata table to an Amazon S3 bucket.

```
$ emrfs diff s3://elasticmapreduce/samples/cloudfront
BOTH | MANIFEST ONLY | S3 ONLY
DIR elasticmapreduce/samples/cloudfront
DIR elasticmapreduce/samples/cloudfront/code/
DIR elasticmapreduce/samples/cloudfront/input/
DIR elasticmapreduce/samples/cloudfront/logprocessor.jar
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-14.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-15.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-16.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-17.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-18.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-19.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-20.WxYz1234
DIR elasticmapreduce/samples/cloudfront/code/cloudfront-loganalyzer.tgz
```

emrfs delete sub-command

[Options] for emrfs delete

Option	Description	Required
-m *METADATA_NAME* | --metadata-name *METADATA_NAME*	*METADATA_NAME* is the name of the DynamoDB metadata table. If the *METADATA_NAME*	No

Option	Description	Required
	argument is not supplied, the default value is `EmrFSMetadata` .	
s3://s3Path	The path to the Amazon S3 bucket you are tracking for consistent view. Buckets sync recursively.	Yes
`-t` *TIME* | `--time` *TIME*	The expiration time (interpreted using the time unit argument). All metadata entries older than the *TIME* argument are deleted for the specified bucket.	
`-u` *UNIT* | `--time-unit` *UNIT*	The measure used to interpret the time argument (nanoseconds, microseconds, milliseconds, seconds, minutes, hours, or days). If no argument is specified, the default value is `days`.	
`--read-consumption` *READ_CONSUMPTION*	The requested amount of available read throughput used for the **delete** operation. If the *READ_CONSUMPTION* argument is not specified, the default value is `400`.	No
`--write-consumption` *WRITE_CONSUMPTION*	The requested amount of available write throughput used for the **delete** operation. If the *WRITE_CONSUMPTION* argument is not specified, the default value is `100`.	No

Example emrfs delete example

The following example removes all objects in an Amazon S3 bucket from the tracking metadata for consistent view.

```
$ emrfs delete s3://elasticmapreduce/samples/cloudfront
entries deleted: 11
```

emrfs import sub-command

[Options] for emrfs import

Option	Description	Required
-m *METADATA_NAME* | --metadata-name *METADATA_NAME*	*METADATA_NAME* is the name of the DynamoDB metadata table. If the *METADATA_NAME* argument is not supplied, the default value is EmrFSMetadata .	No
s3://s3Path	The path to the Amazon S3 bucket you are tracking for consistent view. Buckets sync recursive ly.	Yes
--read-consumption *READ_CONSUMPTION*	The requested amount of available read throughpu t used for the **delete** operation. If the *READ_CONS UMPTION* argument is not specified, the default value is 400.	No
--write-consumption *WRITE_CONSUMPTION*	The requested amount of available write throughput used for the **delete** operation. If the *WRITE_CONSUMPTION* argument is not specified , the default value is 100.	No

Example emrfs import example

The following example imports all objects in an Amazon S3 bucket with the tracking metadata for consistent view. All unknown keys are ignored.

```
$ emrfs import s3://elasticmapreduce/samples/cloudfront
```

emrfs sync sub-command

[Options] for emrfs sync

Option	Description	Required
-m *METADATA_NAME* | --metadata-name *METADATA_NAME*	*METADATA_NAME* is the name of the DynamoDB metadata table. If the *METADATA_NAME* argument is not supplied, the default value is EmrFSMetadata .	No
s3://s3Path	The path to the Amazon S3 bucket you are tracking for consistent view. Buckets sync recursively.	Yes
--read-consumption *READ_CONSUMPTION*	The requested amount of available read throughput used for the **delete** operation. If the *READ_CONSUMPTION* argument is not specified, the default value is 400.	No
--write-consumption *WRITE_CONSUMPTION*	The requested amount of available write throughput used for the **delete** operation. If the *WRITE_CONSUMPTION* argument is not specified, the default value is 100.	No

Example emrfs sync command example

The following example imports all objects in an Amazon S3 bucket with the tracking metadata for consistent view. All unknown keys are deleted.

```
$ emrfs sync s3://elasticmapreduce/samples/cloudfront
Synching samples/cloudfront                                       0 added | 0 updated |
 0 removed | 0 unchanged
Synching samples/cloudfront/code/                                 1 added | 0 updated |
 0 removed | 0 unchanged
Synching samples/cloudfront/                                      2 added | 0 updated |
 0 removed | 0 unchanged
```

```
Synching samples/cloudfront/input/                                9 added | 0 updated |
 0 removed | 0 unchanged
Done synching s3://elasticmapreduce/samples/cloudfront            9 added | 0 updated |
 1 removed | 0 unchanged
creating 3 folder key(s)
folders written: 3
```

emrfs read-sqs sub-command

[Options] for emrfs read-sqs

Option	Description	Required
-q *QUEUE_NAME* | --queue-name *QUEUE_NAME*	*QUEUE_NAME* is the name of the Amazon SQS queue configured in `emrfs-site.xml`. The default value is **`EMRFS-Inconsistency-<jobFlowId>`**.	Yes
-o *OUTPUT_FILE* | --output-file *OUTPUT_FILE*	*OUTPUT_FILE* is the path to the output file on the master node's local file system. Messages read from the queue are written to this file.	Yes

emrfs delete-sqs sub-command

[Options] for emrfs delete-sqs

Option	Description	Required
-q *QUEUE_NAME* | --queue-name *QUEUE_NAME*	*QUEUE_NAME* is the name of the Amazon SQS queue configured in `emrfs-site.xml`. The default value is **`EMRFS-Inconsistency-<jobFlowId>`**.	Yes

Submitting EMRFS CLI commands as steps

The following example shows how to use the `emrfs` utility on the master node by leveraging the AWS CLI or API and the `command-runner.jar` to run the `emrfs` command as a step. The example uses the AWS SDK for Python (Boto3) to add a step to a cluster which adds objects in an Amazon S3 bucket to the default EMRFS metadata table.

```
import boto3
from botocore.exceptions import ClientError


def add_emrfs_step(command, bucket_url, cluster_id, emr_client):
    """
    Add an EMRFS command as a job flow step to an existing cluster.

    :param command: The EMRFS command to run.
    :param bucket_url: The URL of a bucket that contains tracking metadata.
    :param cluster_id: The ID of the cluster to update.
    :param emr_client: The Boto3 Amazon EMR client object.
    :return: The ID of the added job flow step. Status can be tracked by calling
             the emr_client.describe_step() function.
    """
    job_flow_step = {
        "Name": "Example EMRFS Command Step",
        "ActionOnFailure": "CONTINUE",
        "HadoopJarStep": {
            "Jar": "command-runner.jar",
            "Args": ["/usr/bin/emrfs", command, bucket_url],
        },
    }

    try:
        response = emr_client.add_job_flow_steps(
            JobFlowId=cluster_id, Steps=[job_flow_step]
        )
        step_id = response["StepIds"][0]
        print(f"Added step {step_id} to cluster {cluster_id}.")
    except ClientError:
        print(f"Couldn't add a step to cluster {cluster_id}.")
        raise
    else:
        return step_id


def usage_demo():
    emr_client = boto3.client("emr")
    # Assumes the first waiting cluster has EMRFS enabled and has created metadata
```

```
    # with the default name of 'EmrFSMetadata'.
    cluster = emr_client.list_clusters(ClusterStates=["WAITING"])["Clusters"][0]
    add_emrfs_step(
        "sync", "s3://elasticmapreduce/samples/cloudfront", cluster["Id"], emr_client
    )


if __name__ == "__main__":
    usage_demo()
```

You can use the `step_id` value returned to check the logs for the result of the operation.

Authorizing access to EMRFS data in Amazon S3

By default, the EMR role for EC2 determines the permissions for accessing EMRFS data in Amazon S3. The IAM policies that are attached to this role apply regardless of the user or group making the request through EMRFS. The default is `EMR_EC2_DefaultRole`. For more information, see Service role for cluster EC2 instances (EC2 instance profile).

Beginning with Amazon EMR release version 5.10.0, you can use a security configuration to specify IAM roles for EMRFS. This allows you to customize permissions for EMRFS requests to Amazon S3 for clusters that have multiple users. You can specify different IAM roles for different users and groups, and for different Amazon S3 bucket locations based on the prefix in Amazon S3. When EMRFS makes a request to Amazon S3 that matches users, groups, or the locations that you specify, the cluster uses the corresponding role that you specify instead of the EMR role for EC2. For more information, see Configure IAM roles for EMRFS requests to Amazon S3.

Alternatively, if your Amazon EMR solution has demands beyond what IAM roles for EMRFS provides, you can define a custom credentials provider class, which allows you to customize access to EMRFS data in Amazon S3.

Creating a custom credentials provider for EMRFS data in Amazon S3

To create a custom credentials provider, you implement the AWSCredentialsProvider and the Hadoop Configurable classes.

For a detailed explanation of this approach, see Securely analyze data from another AWS account with EMRFS in the AWS Big Data blog. The blog post includes a tutorial that walks you through the process end-to-end, from creating IAM roles to launching the cluster. It also provides a Java code example that implements the custom credential provider class.

The basic steps are as follows:

To specify a custom credentials provider

1. Create a custom credentials provider class compiled as a JAR file.
2. Run a script as a bootstrap action to copy the custom credentials provider JAR file to the `/usr/share/aws/emr/emrfs/auxlib` location on the cluster's master node. For more information about bootstrap actions, see [(Optional) Create bootstrap actions to install additional software](#).
3. Customize the `emrfs-site` classification to specify the class that you implement in the JAR file. For more information about specifying configuration objects to customize applications, see [Configuring applications](#) in the *Amazon EMR Release Guide*.

   The following example demonstrates a `create-cluster` command that launches a Hive cluster with common configuration parameters, and also includes:

   - A bootstrap action that runs the script, *`copy_jar_file.sh`*, which is saved to *`mybucket`* in Amazon S3.
   - An `emrfs-site` classification that specifies a custom credentials provider defined in the JAR file as *`MyCustomCredentialsProvider`*

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --applications Name=Hive \
--bootstrap-actions '[{"Path":"s3://mybucket/copy_jar_file.sh","Name":"Custom
 action"}]' \
--ec2-attributes '{"KeyName":"MyKeyPair","InstanceProfile":"EMR_EC2_DefaultRole",\
"SubnetId":"subnet-xxxxxxxx","EmrManagedSlaveSecurityGroup":"sg-xxxxxxxx",\
"EmrManagedMasterSecurityGroup":"sg-xxxxxxxx"}' \
--service-role EMR_DefaultRole_V2 --enable-debugging --release-label emr-7.2.0 \
--log-uri 's3n://my-emr-log-bucket/' --name 'test-awscredentialsprovider-emrfs' \
--instance-type=m5.xlarge --instance-count 3  \
--configurations '[{"Classification":"emrfs-site",\
```

```
"Properties":
{"fs.s3.customAWSCredentialsProvider":"MyAWSCredentialsProviderWithUri"},\
"Configurations":[]}]'
```

Managing the default AWS Security Token Service endpoint

EMRFS uses the AWS Security Token Service (STS) to retrieve temporary security credentials in order to access your AWS resources. Earlier Amazon EMR release versions send all AWS STS requests to a single global endpoint at `https://sts.amazonaws.com`. Amazon EMR release versions 5.31.0 and 6.1.0 and later make requests to Regional AWS STS endpoints instead. This reduces latency and improves session token validity. For more information about AWS STS endpoints, see [Managing AWS STS in an AWS Region](#) in the *AWS Identity and Access Management User Guide*.

When you use Amazon EMR release versions 5.31.0 and 6.1.0 and later, you can override the default AWS STS endpoint. To do so, you must change the `fs.s3.sts.endpoint` property in your `emrfs-site` configuration.

The following AWS CLI example sets the default AWS STS endpoint used by EMRFS to the global endpoint.

```
aws emr create-cluster --release-label <emr-5.33.0> --instance-type m5.xlarge \
--emrfs Args=[fs.s3.sts.endpoint=https://sts.amazonaws.com]
```

Note

Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

Alternatively, you can create a JSON configuration file using the following example, and specify it using the `--configurations` argument of `emr create-cluster`. For more information about using `--configurations`, see the [*AWS CLI Command Reference.*](#)

```
[
  {
    "classification": "emrfs-site",
    "properties": {
      "fs.s3.sts.endpoint": "https://sts.amazonaws.com"
```

```
    }
  }
]
```

Specifying Amazon S3 encryption using EMRFS properties

Important

Beginning with Amazon EMR release version 4.8.0, you can use security configurations to apply encryption settings more easily and with more options. We recommend using security configurations. For information, see Configure data encryption. The console instructions described in this section are available for release versions earlier than 4.8.0. If you use the AWS CLI to configure Amazon S3 encryption both in the cluster configuration and in a security configuration in subsequent versions, the security configuration overrides the cluster configuration.

When you create a cluster, you can specify server-side encryption (SSE) or client-side encryption (CSE) for EMRFS data in Amazon S3 using the console or using `emrfs-site` classification properties through the AWS CLI or EMR SDK. Amazon S3 SSE and CSE are mutually exclusive; you can choose either but not both.

For AWS CLI instructions, see the appropriate section for your encryption type below.

To specify EMRFS encryption options using the AWS Management Console

1. Navigate to the new Amazon EMR console and select **Switch to the old console** from the side navigation. For more information on what to expect when you switch to the old console, see Using the old console.
2. Choose **Create cluster**, **Go to advanced options**.
3. Choose a **Release** of 4.7.2 or earlier.
4. Choose other options for **Software and Steps** as appropriate for your application, and then choose **Next**.
5. Choose settings in the **Hardware** and **General Cluster Settings** panes as appropriate for your application.
6. On the **Security** pane, under **Authentication and encryption**, select the **S3 Encryption (with EMRFS)** option to use.

> **Note**
>
> **S3 server-side encryption with KMS Key Management** (SSE-KMS) is not available when using Amazon EMR release version 4.4 or earlier.

- If you choose an option that uses **AWS Key Management**, choose an **AWS KMS Key ID**. For more information, see Using AWS KMS keys for EMRFS encryption.
- If you choose **S3 client-side encryption with custom materials provider**, provide the **Class name** and the **JAR location**. For more information, see Amazon S3 client-side encryption.

7. Choose other options as appropriate for your application and then choose **Create Cluster**.

Using AWS KMS keys for EMRFS encryption

The AWS KMS encryption key must be created in the same Region as your Amazon EMR cluster instance and the Amazon S3 buckets used with EMRFS. If the key that you specify is in a different account from the one that you use to configure a cluster, you must specify the key using its ARN.

The role for the Amazon EC2 instance profile must have permissions to use the KMS key you specify. The default role for the instance profile in Amazon EMR is `EMR_EC2_DefaultRole`. If you use a different role for the instance profile, or you use IAM roles for EMRFS requests to Amazon S3, make sure that each role is added as a key user as appropriate. This gives the role permissions to use the KMS key. For more information, see Using Key Policies in the *AWS Key Management Service Developer Guide* and Configure IAM roles for EMRFS requests to Amazon S3.

You can use the AWS Management Console to add your instance profile or EC2 instance profile to the list of key users for the specified KMS key, or you can use the AWS CLI or an AWS SDK to attach an appropriate key policy.

Note that Amazon EMR supports only symmetric KMS keys. You cannot use an asymmetric KMS key to encrypt data at rest in an Amazon EMR cluster. For help determining whether a KMS key is symmetric or asymmetric, see Identifying symmetric and asymmetric KMS keys.

The procedure below describes how to add the default Amazon EMR instance profile, `EMR_EC2_DefaultRole` as a *key user* using the AWS Management Console. It assumes that you have already created a KMS key. To create a new KMS key, see Creating Keys in the *AWS Key Management Service Developer Guide*.

To add the EC2 instance profile for Amazon EMR to the list of encryption key users

1. Sign in to the AWS Management Console and open the AWS Key Management Service (AWS KMS) console at https://console.aws.amazon.com/kms.
2. To change the AWS Region, use the Region selector in the upper-right corner of the page.
3. Select the alias of the KMS key to modify.
4. On the key details page under **Key Users**, choose **Add**.
5. In the **Add key users** dialog box, select the appropriate role. The name of the default role is `EMR_EC2_DefaultRole`.
6. Choose **Add**.

Amazon S3 server-side encryption

When you set up Amazon S3 server-side encryption, Amazon S3 encrypts data at the object level as it writes the data to disk and decrypts the data when it is accessed. For more information about SSE, see Protecting data using server-side encryption in the *Amazon Simple Storage Service User Guide*.

You can choose between two different key management systems when you specify SSE in Amazon EMR:

- **SSE-S3** – Amazon S3 manages keys for you.
- **SSE-KMS** – You use an AWS KMS key to set up with policies suitable for Amazon EMR. For more information about key requirements for Amazon EMR, see Using AWS KMS keys for encryption.

SSE with customer-provided keys (SSE-C) is not available for use with Amazon EMR.

To create a cluster with SSE-S3 enabled using the AWS CLI

- Type the following command:

```
aws emr create-cluster --release-label emr-4.7.2 or earlier \
--instance-count 3 --instance-type m5.xlarge --emrfs Encryption=ServerSide
```

You can also enable SSE-S3 by setting the fs.s3.enableServerSideEncryption property to true in `emrfs-site` properties. See the example for SSE-KMS below and omit the property for Key ID.

To create a cluster with SSE-KMS enabled using the AWS CLI

> **Note**
>
> SSE-KMS is available only in Amazon EMR release version 4.5.0 and later.

- Type the following AWS CLI command to create a cluster with SSE-KMS, where *keyID* is an AWS KMS key, for example, *a4567b8-9900-12ab-1234-123a45678901*:

  ```
  aws emr create-cluster --release-label emr-4.7.2 or earlier --instance-count 3 \
  --instance-type m5.xlarge --use-default-roles \
  --emrfs Encryption=ServerSide,Args=[fs.s3.serverSideEncryption.kms.keyId=keyId]
  ```

  --OR--

  Type the following AWS CLI command using the `emrfs-site` classification and provide a configuration JSON file with contents as shown similar to `myConfig.json` in the example below:

  ```
  aws emr create-cluster --release-label emr-4.7.2 or earlier --instance-count 3
   --instance-type m5.xlarge --applications Name=Hadoop --configurations file://
  myConfig.json --use-default-roles
  ```

  Example contents of **myConfig.json**:

  ```
  [
    {
      "Classification":"emrfs-site",
      "Properties": {
         "fs.s3.enableServerSideEncryption": "true",
         "fs.s3.serverSideEncryption.kms.keyId":"a4567b8-9900-12ab-1234-123a45678901"
      }
    }
  ]
  ```

Configuration properties for SSE-S3 and SSE-KMS

These properties can be configured using the `emrfs-site` configuration classification. SSE-KMS is available only in Amazon EMR release version 4.5.0 and later.

Property	Default value	Description
`fs.s3.enableServerSideEncryption`	**`false`**	When set to **`true`**, objects stored in Amazon S3 are encrypted using server-side encryption. If no key is specified, SSE-S3 is used.
`fs.s3.serverSideEncryption.kms.keyId`	**`n/a`**	Specifies an AWS KMS key ID or ARN. If a key is specified, SSE-KMS is used.

Amazon S3 client-side encryption

With Amazon S3 client-side encryption, the Amazon S3 encryption and decryption takes place in the EMRFS client on your cluster. Objects are encrypted before being uploaded to Amazon S3 and decrypted after they are downloaded. The provider you specify supplies the encryption key that the client uses. The client can use keys provided by AWS KMS (CSE-KMS) or a custom Java class that provides the client-side root key (CSE-C). The encryption specifics are slightly different between CSE-KMS and CSE-C, depending on the specified provider and the metadata of the object being decrypted or encrypted. For more information about these differences, see [Protecting data using client-side encryption] in the *Amazon Simple Storage Service User Guide*.

> **Note**
>
> Amazon S3 CSE only ensures that EMRFS data exchanged with Amazon S3 is encrypted; not all data on cluster instance volumes is encrypted. Furthermore, because Hue does not use EMRFS, objects that the Hue S3 File Browser writes to Amazon S3 are not encrypted.

To specify CSE-KMS for EMRFS data in Amazon S3 using the AWS CLI

- Type the following command and replace *MyKMSKeyID* with the Key ID or ARN of the KMS key to use:

```
aws emr create-cluster --release-label emr-4.7.2 or earlier
--emrfs Encryption=ClientSide,ProviderType=KMS,KMSKeyId=MyKMSKeyId
```

Creating a custom key provider

Depending on the type of encryption you use when creating a custom key provider, the application must also implement different EncryptionMaterialsProvider interfaces. Both interfaces are available in the AWS SDK for Java version 1.11.0 and later.

- To implement Amazon S3 encryption, use the [com.amazonaws.services.s3.model.EncryptionMaterialsProvider interface](https://docs.aws.amazon.com/AWSJavaSDK/latest/javadoc/com/amazonaws/services/s3/model/EncryptionMaterialsProvider.html).
- To implement local disk encryption, use the com.amazonaws.services.elasticmapreduce.spi.security.EncryptionMaterialsProvider interface.

You can use any strategy to provide encryption materials for the implementation. For example, you might choose to provide static encryption materials or integrate with a more complex key management system.

If you're using Amazon S3 encryption, you must use the encryption algorithms **AES/GCM/NoPadding** for custom encryption materials.

If you're using local disk encryption, the encryption algorithm to use for custom encryption materials varies by EMR release. For Amazon EMR 7.0.0 and lower, you must use **AES/GCM/NoPadding**. For Amazon EMR 7.1.0 and higher, you must use **AES**.

The EncryptionMaterialsProvider class gets encryption materials by encryption context. Amazon EMR populates encryption context information at runtime to help the caller determine the correct encryption materials to return.

Example Example: Using a custom key provider for Amazon S3 encryption with EMRFS

When Amazon EMR fetches the encryption materials from the EncryptionMaterialsProvider class to perform encryption, EMRFS optionally populates the materialsDescription argument with two fields: the Amazon S3 URI for the object and the JobFlowId of the cluster, which can be used by the EncryptionMaterialsProvider class to return encryption materials selectively.

For example, the provider may return different keys for different Amazon S3 URI prefixes. It is the description of the returned encryption materials that is eventually stored with the Amazon S3 object rather than the materialsDescription value that is generated by EMRFS and passed to the provider. While decrypting an Amazon S3 object, the encryption materials description is passed to the EncryptionMaterialsProvider class, so that it can, again, selectively return the matching key to decrypt the object.

An EncryptionMaterialsProvider reference implementation is provided below. Another custom provider, EMRFSRSAEncryptionMaterialsProvider, is available from GitHub.

```
import com.amazonaws.services.s3.model.EncryptionMaterials;
import com.amazonaws.services.s3.model.EncryptionMaterialsProvider;
import com.amazonaws.services.s3.model.KMSEncryptionMaterials;
import org.apache.hadoop.conf.Configurable;
import org.apache.hadoop.conf.Configuration;

import java.util.Map;

/**
 * Provides KMSEncryptionMaterials according to Configuration
 */
public class MyEncryptionMaterialsProviders implements EncryptionMaterialsProvider,
 Configurable{
  private Configuration conf;
  private String kmsKeyId;
  private EncryptionMaterials encryptionMaterials;

  private void init() {
    this.kmsKeyId = conf.get("my.kms.key.id");
    this.encryptionMaterials = new KMSEncryptionMaterials(kmsKeyId);
  }

  @Override
  public void setConf(Configuration conf) {
    this.conf = conf;
    init();
  }

  @Override
```

```
  public Configuration getConf() {
    return this.conf;
  }

  @Override
  public void refresh() {

  }

  @Override
  public EncryptionMaterials getEncryptionMaterials(Map<String, String>
 materialsDescription) {
    return this.encryptionMaterials;
  }

  @Override
  public EncryptionMaterials getEncryptionMaterials() {
    return this.encryptionMaterials;
  }
}
```

Specifying a custom materials provider using the AWS CLI

To use the AWS CLI, pass the `Encryption`, `ProviderType`, `CustomProviderClass`, and `CustomProviderLocation` arguments to the `emrfs` option.

```
aws emr create-cluster --instance-type m5.xlarge --release-label emr-4.7.2 or earlier
 --emrfs Encryption=ClientSide,ProviderType=Custom,CustomProviderLocation=s3://
mybucket/myfolder/provider.jar,CustomProviderClass=classname
```

Setting `Encryption` to `ClientSide` enables client-side encryption, `CustomProviderClass` is the name of your `EncryptionMaterialsProvider` object, and `CustomProviderLocation` is the local or Amazon S3 location from which Amazon EMR copies `CustomProviderClass` to each node in the cluster and places it in the classpath.

Specifying a custom materials provider using an SDK

To use an SDK, you can set the property `fs.s3.cse.encryptionMaterialsProvider.uri` to download the custom `EncryptionMaterialsProvider` class that you store in Amazon S3 to each node in your cluster. You configure this in `emrfs-site.xml` file along with CSE enabled and the proper location of the custom provider.

For example, in the AWS SDK for Java using RunJobFlowRequest, your code might look like the following:

```
<snip>
  Map<String,String> emrfsProperties = new HashMap<String,String>();
      emrfsProperties.put("fs.s3.cse.encryptionMaterialsProvider.uri","s3://mybucket/
MyCustomEncryptionMaterialsProvider.jar");
      emrfsProperties.put("fs.s3.cse.enabled","true");
      emrfsProperties.put("fs.s3.consistent","true");

 emrfsProperties.put("fs.s3.cse.encryptionMaterialsProvider","full.class.name.of.EncryptionMate

  Configuration myEmrfsConfig = new Configuration()
      .withClassification("emrfs-site")
      .withProperties(emrfsProperties);

  RunJobFlowRequest request = new RunJobFlowRequest()
   .withName("Custom EncryptionMaterialsProvider")
   .withReleaseLabel("emr-7.2.0")
   .withApplications(myApp)
   .withConfigurations(myEmrfsConfig)
   .withServiceRole("EMR_DefaultRole_V2")
   .withJobFlowRole("EMR_EC2_DefaultRole")
   .withLogUri("s3://myLogUri/")
   .withInstances(new JobFlowInstancesConfig()
    .withEc2KeyName("myEc2Key")
    .withInstanceCount(2)
    .withKeepJobFlowAliveWhenNoSteps(true)
    .withMasterInstanceType("m5.xlarge")
    .withSlaveInstanceType("m5.xlarge")
   );

  RunJobFlowResult result = emr.runJobFlow(request);
</snip>
```

Custom EncryptionMaterialsProvider with arguments

You may need to pass arguments directly to the provider. To do this, you can use the `emrfs-site` configuration classification with custom arguments defined as properties. An example configuration is shown below, which is saved as a file, `myConfig.json`:

```
[
```

```
    {
      "Classification": "emrfs-site",
      "Properties": {
        "myProvider.arg1":"value1",
     "myProvider.arg2":"value2"
      }
    }
 ]
```

Using the `create-cluster` command from the AWS CLI, you can use the `--configurations` option to specify the file as shown below:

```
aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge
 --instance-count 2 --configurations file://myConfig.json --emrfs
 Encryption=ClientSide,CustomProviderLocation=s3://mybucket/myfolder/
myprovider.jar,CustomProviderClass=classname
```

Configuring EMRFS S3EC V2 support

S3 Java SDK releases (1.11.837 and later) support encryption client Version 2 (S3EC V2) with various security enhancements. For more information, see the S3 blog post Updates to the Amazon S3 encryption client. Also, refer to Amazon S3 encryption client migration in the AWS SDK for Java Developer Guide.

Encryption client V1 is still available in the SDK for backward compatibility. By default EMRFS will use S3EC V1 to encrypt and decrypt S3 objects if CSE is enabled.

S3 objects encrypted with S3EC V2 cannot be decrypted by EMRFS on an EMR cluster whose release version is earlier than emr-5.31.0 (emr-5.30.1 and earlier, emr-6.1.0 and earlier).

Example Configure EMRFS to use S3EC V2

To configure EMRFS to use S3EC V2, add the following configuration:

```
{
  "Classification": "emrfs-site",
  "Properties": {
    "fs.s3.cse.encryptionV2.enabled": "true"
  }
}
```

`emrfs-site.xml` Properties for Amazon S3 client-side encryption

Property	Default value	Description
`fs.s3.cse.enabled`	**`false`**	When set to **`true`**, EMRFS objects stored in Amazon S3 are encrypted using client-side encryption.
`fs.s3.cse.encryptionV2.enabled`	**`false`**	When set to `true`, EMRFS uses S3 encryption client Version 2 to encrypt and decrypt objects on S3. Available for EMR version 5.31.0 and later.
`fs.s3.cse.encryptionMaterialsProvider.uri`	**N/A**	Applies when using custom encryption materials. The Amazon S3 URI where the JAR with the `EncryptionMaterialsProvider` is located. When you provide this URI, Amazon EMR automatically downloads the JAR to all nodes in the cluster.
`fs.s3.cse.encryptionMaterialsProvider`	**N/A**	The `EncryptionMaterialsProvider` class path used with client-side encryption. When using CSE-KMS, specify `com.amazon.ws.emr.hadoop.fs.cse.KMSEncryptionMaterialsProvider`.

Property	Default value	Description
fs.s3.cse.materialsDescription.enabled	**false**	When set to true, populates the materialsDescription of encrypted objects with the Amazon S3 URI for the object and the JobFlowId. Set to true when using custom encryption materials.
fs.s3.cse.kms.keyId	**N/A**	Applies when using CSE-KMS. The value of the KeyId, ARN, or alias of the KMS key used for encryption.
fs.s3.cse.cryptoStorageMode	**ObjectMetadata**	The Amazon S3 storage mode. By default, the description of the encryption information is stored in the object metadata. You can also store the description in an instruction file. Valid values are ObjectMetadata and InstructionFile. For more information, see Client-side data encryption with the AWS SDK for Java and Amazon S3.

Amazon CloudWatch agent

The Amazon CloudWatch agent on Amazon EMR is a tool that can monitor the Amazon EC2 instances in your EMR cluster. You can store and view the metrics that you collect with the CloudWatch agent in CloudWatch. For more information about the CloudWatch agent, see the *Amazon CloudWatch User Guide*.

> **Note**
>
> You incur additional charges if you use other AWS services to publish, query, or view Amazon CloudWatch agent metrics. See the following pages for more pricing information.
>
> - Amazon CloudWatch Pricing
> - Amazon Managed Service for Prometheus Pricing
> - Amazon Managed Grafana Pricing
>
> These metrics are separate from the free metrics that Amazon EMR provides under the `AWS/ElasticMapReduce` namespace in CloudWatch. For more information about the metrics that the CloudWatch agent doesn't publish, see Monitoring Amazon EMR metrics with CloudWatch.

With Amazon EMR 7.0 and higher, you can install a custom version of the Amazon CloudWatch agent on your cluster to collect metrics from your EMR cluster. The agent collects the metrics from all nodes in the cluster, gathers them on the primary node, and publishes the metrics from that node to the cloud.

You can install the agent when you create a new cluster in the console, or when you use the `create-cluster` API. For more information, see Create an EMR cluster that uses Amazon CloudWatch agent. By default, an Amazon EMR cluster that runs the release 7.x series publishes 34 system-level metrics to CloudWatch at 60-second intervals, but you can configure the agent to publish different metrics. Another option is to publish metrics to Amazon Managed Service for Prometheus, and you can choose which metrics to publish as well. For different use cases and setups, you might configure the agent to view and query the metrics in the CloudWatch console, Amazon Managed Grafana, or through the APIs for CloudWatch or Amazon Managed Service for Prometheus. These AWS services incur charges when you use them to store and query metrics.

The CloudWatch agent on Amazon EMR can publish system metrics as well as JMX metrics for the following services on your Amazon EMR cluster.

- Hadoop DataNode
- Hadoop NameNode
- Yarn NodeManager
- Yarn ResourceManager
- HBase Master – Amazon EMR 7.1+ only
- HBase RegionServer – Amazon EMR 7.1+ only
- HBase ThriftServer – Amazon EMR 7.1+ only

For more information about available metrics and how to configure the CloudWatch agent on Amazon EMR, see Configuring CloudWatch agent for Amazon EMR.

The following table lists the version of AmazonCloudWatchAgent included in the latest release of the Amazon EMR 7.x series, along with the components that Amazon EMR installs with AmazonCloudWatchAgent.

For the version of components installed with AmazonCloudWatchAgent in this release, see Release 7.2.0 Component Versions.

AmazonCloudWatchAgent version information for emr-7.2.0

Amazon EMR Release Label	AmazonCloudWatchAgent Version	Components Installed With AmazonCloudWatchAgent
emr-7.2.0	AmazonCloudWatchAgent 1.300032.2	adot-java-agent, emrfs, emr-amazon-cloudwatch-agent , emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-ya

Amazon EMR Release Label	AmazonCloudWatchAgent Version	Components Installed With AmazonCloudWatchAgent
		rn-resourcemanager, hadoop-yarn-timeline-server

Topics

- Create an EMR cluster that uses Amazon CloudWatch agent
- Default metrics for CloudWatch agent with Amazon EMR
- Configuring CloudWatch agent for Amazon EMR
- Considerations and limitations
- CloudWatch agent release history

Create an EMR cluster that uses Amazon CloudWatch agent

The procedures in this section describe the steps to create a cluster in Amazon EMR with Amazon CloudWatch agent from the AWS Management Console and the AWS CLI.

Topics

- Required IAM permissions for CloudWatch agent
- Required CloudWatch agent endpoint
- Create an EMR cluster

Required IAM permissions for CloudWatch agent

The CloudWatch agent requires the AWS Identity and Access Management (IAM) `cloudwatch:PutMetricData` permission in the Amazon EC2 instance profile for Amazon EMR. The Amazon EMR default role already has this permission. You can create the default role from the AWS CLI with `aws emr create-default-roles`. For more information, see Service role for cluster EC2 instances (EC2 instance profile) in the *Amazon EMR Management Guide*.

The following example IAM policy includes the `cloudwatch:PutMetricData` permission:

```
{
```

```
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "cloudwatch:PutMetricData",
      "Resource": "*"
    }
  ]
}
```

Required CloudWatch agent endpoint

To publish metrics to CloudWatch for an EMR cluster in a private subnet, create a CloudWatch agent endpoint and associate with the VPC that the private subnet is in.

For more information about the CloudWatch endpoints for each AWS Region, see Amazon CloudWatch endpoints and quotas in the *AWS General Reference Guide*.

Create an EMR cluster

Once you have set up the required permissions and endpoint for use with the CloudWatch agent, use the AWS Management Console or the AWS CLI to create a new cluster with the agent installed.

Console

To create a cluster with Amazon CloudWatch agent from the console

1. Navigate to the Amazon EMR console.
2. Choose **Create cluster**.
3. Under **Name and applications**, choose an Amazon EMR release of 7.0.0 or higher.
4. Under **Application bundle**, select the bundle or apps that you want to install to your cluster, and include **CloudWatch agent** with your selections.
5. Proceed to create the cluster to serve your use case needs.

AWS CLI

In the AWS CLI, you can add Amazon CloudWatch agent to a cluster with the `--applications` parameter for `create-cluster`.

To create a cluster with Amazon CloudWatch agent from the AWS CLI

- When you create a cluster, use a command similar to the following to include the Amazon CloudWatch agent. Replace *myKey* with the name of your EC2 key pair.

```
aws emr create-cluster --name "Spark cluster with CloudWatch agent" \
--release-label emr-7.0.0 \
--applications Name=Spark Name=AmazonCloudWatchAgent \
--ec2-attributes KeyName=myKey --instance-type m7g.2xlarge \
--instance-count 3 --use-default-roles
```

For more details on how to use Amazon EMR with the AWS CLI, see the AWS CLI Command Reference.

Default metrics for CloudWatch agent with Amazon EMR

When you install the Amazon CloudWatch agent on Amazon EMR, the default configuration publishes the following system metrics for all of the instances in your cluster unless you configure the agent differently. For definitions of each metric, see Metrics collected by the CloudWatch agent in the *Amazon CloudWatch User Guide*.

CPU

CPU metrics

- `cpu_usage_active`
- `cpu_usage_guest`
- `cpu_usage_guest_nice`
- `cpu_usage_idle`
- `cpu_usage_iowait`
- `cpu_usage_irq`
- `cpu_usage_nice`
- `cpu_usage_softirq`
- `cpu_usage_steal`
- `cpu_usage_system`

- `cpu_usage_user`

Disk

Disk metrics

- `disk_free`
- `disk_total`
- `disk_used`
- `disk_used_percent`

Memory

Memory metrics

- `mem_active`
- `mem_available`
- `mem_available_percent`
- `mem_free`
- `mem_inactive`
- `mem_total`
- `mem_used`
- `mem_used_percent`
- `mem_buffered`
- `mem_cached`

Network IO

Network IO metrics

- `net_bytes_recv`
- `net_bytes_sent`
- `net_packets_recv`
- `net_packets_sent`

Process

Process metrics

- `processes_running`
- `processes_total`

Swap

Swap metrics

- `swap_free`
- `swap_used`
- `swap_used_percent`

By default, the agent publishes all metrics to CloudWatch under the custom namespace `CWAgent`, and under the schema `cluster.id`, `instance.id`, `node.type`, `service.name`. Use the following steps to view these default metrics that the CloudWatch agent publishes for Amazon EMR:

1. Navigate to the Amazon CloudWatch console.
2. Choose the **Metrics** tab, and then **All metrics**.
3. Under **Custom namespaces**, select **`CWAgent`**. Then, select the schema: **`cluster.id`, `instance.id`, `node.type`, `service.name`**.
4. Continue to query the metrics from the CloudWatch interface as your use case requires.

Amazon EMR attaches the following labels to each metric: `["cluster.id", "instance.id', "node.type", "service.name"]`. Consider the following with regard to these labels:

- The value of the `cluster.id` label is the ID of the EMR cluster that produced the metric. An example value for the `cluster.id` label is: `j-123456789ABC`.
- The value of the `instance.id` label is the ID of the instance in the EMR cluster that produced the metric. An example value for the `instance.id` label is `i-01bcf5f140f3355777`.
- The value of the `node.type` label represents the node type of the instance in the cluster that produced the metric. This value can be either `master`, `core`, or `task`.

- The value of the `service.name` label is the name of the service that produced the metric. The default `service.name` value for the default system metrics is `system`.

> **Note**
>
> Amazon EMR release 7.0.0 uses a slightly different metric label schema. `cluster.id` is `jobflow.id`, and `node.type` is unavailable. `instance.id` and `service.name` are available.

Configuring CloudWatch agent for Amazon EMR

Amazon EMR 7.0.0 and higher include the Amazon CloudWatch agent. The following sections describe how to configure the agent on your EMR cluster as described in Create an EMR cluster that uses Amazon CloudWatch agent.

Topics

- Configure CloudWatch agent for Amazon EMR 7.1.0
- Configure CloudWatch agent for Amazon EMR 7.0.0

Configure CloudWatch agent for Amazon EMR 7.1.0

Starting with Amazon EMR 7.1.0, you can configure the Amazon CloudWatch agent to use additional system metrics, add application metrics, and change metrics destination by using the Amazon EMR configuration API. For more information about how to use the EMR configuration API to configure your cluster's applications, see Configure applications.

> **Note**
>
> 7.1.0 only supports the reconfiguration type `OVERWRITE`. For more information about the reconfiguration types, see Considerations when you reconfigure an instance group.

Topics

- Configuration schema
- System metrics configurations examples

- Application metrics configurations examples
- Amazon Managed Service for Prometheus example

Configuration schema

`emr-metrics` has the following classifications:

- `emr-system-metrics` — configure system metrics, such as CPU, disk, and memory.
- `emr-hadoop-hdfs-datanode-metrics` — configure Hadoop DataNode JMX metrics
- `emr-hadoop-hdfs-namenode-metrics` — configure Hadoop NameNode JMX metrics
- `emr-hadoop-yarn-nodemanager-metrics` — configure Yarn NodeManager JMX metrics
- `emr-hadoop-yarn-resourcemanager-metrics` — configure Yarn ResourceManager JMX metrics
- `emr-hbase-master-metrics` — configure HBase Master JMX metrics
- `emr-hbase-region-server-metrics` — configure HBase Region Server JMX metrics
- `emr-hbase-rest-server-metrics` — configure HBase REST Server JMX metrics
- `emr-hbase-thrift-server-metrics` — configure HBase Thrift Server JMX metrics

The following tables describe the available properties and configurations for all of the classifications.

emr-metrics properties

Property	Required	Description	Default value	Possible values	Notes
`metrics_destination`	Optional	Determines whether cluster metrics are published to Amazon CloudWatch or Amazon Managed	"CLOUDWATCH"	"CLOUDWATCH", "PROMETHEUS"	This property is case-insensitive. For example, "Cloudwatch" is the same as "CLOUDWATCH".

Property	Required	Description	Default value	Possible values	Notes
		Service for Prometheus.			
`prometheus_endpoint`	Optional	If `metrics_destination` is set to "PROMETHEUS", this property configures the CloudWatch agent to send metrics to the provided Amazon Managed Service for Prometheus remote write endpoint.	N/A	Any valid Amazon Managed Service for Prometheus remote write URL. The remote write URL format is `https://aps-workspaces.<region>.amazonaws.com/workspaces/<workspace_id>/api/v1/remote_write`	This field is required if `metrics_destination` is set to "PROMETHEUS". Provisioning will fail if you don't provide a key or if the value is an empty string.

emr-system-metrics properties

Property	Required	Description	Default value	Possible values	Notes
`metrics_collection_interval`	Optional	How often in seconds metrics are collected and	"60"	A string specifying the number of seconds.	You can override this property with the

Property	Required	Description	Default value	Possible values	Notes
		published from the CloudWatch agent.		Only accepts whole numbers.	`metrics_collection_interval` property from individual metric groups.

emr-system-metrics configurations

cpu

Property	Required	Description	Default value	Possible values	Notes
`metrics`	Optional	The list of CPU metrics for the agent to collect.	See Default metrics for CloudWatch agent with Amazon EMR	A comma-separated list of valid CPU metric names with or without the `cpu_` prefix, such as `usage_active` and `cpu_time_idle`. See Metrics collected by the CloudWatc	Specifying an empty string means to not publish any CPU metrics.

Property	Required	Description	Default value	Possible values	Notes
				h agent for valid metrics.	
`metrics_collection_interval`	Optional	How often in seconds the agent should collect and publish CPU metrics.	The value of the global `metrics_collection_interval` .	A string specifying the number of seconds. Accepts only whole numbers.	This value overrides the global `metrics_collection_interval` property only for CPU metrics.
`drop_original_metrics`	Optional	List of CPU metrics for which to not publish unaggregated metrics.	No unaggregated CPU metrics published.	A comma-separated list of CPU metrics that are also specified in the metrics property. An empty string means to publish all CPU metrics.	The CloudWatch agent aggregates all metrics by cluster ID, instance ID, node type, and service name. By default, the CloudWatch agent doesn't publish the per-resource metrics for metrics with multiple resources.

Property	Required	Description	Default value	Possible values	Notes
`resources`	Optional	Determines whether the agent will publish per-core metrics.	"*"	"*" enable per-core metrics. "" disable per-core metrics.	The CloudWatch agent only publishes per-core metrics for CPU metrics that aren't dropped in `drop_original_metrics` .

disk

Property	Required	Description	Default value	Possible values	Notes
`metrics`	Optional	The list of disk metrics for the agent to collect.	See [Default metrics for CloudWatch agent with Amazon EMR](#)	A comma-separated list of valid disk metric names with or without the `disk_` prefix, such as `disk_total` and `used_percent` . See [Metrics collected](#)	Specifying an empty string means to not publish any disk metrics.

Property	Required	Description	Default value	Possible values	Notes
				by the CloudWatch agent for valid metrics.	
`metrics_collection_interval`	Optional	How often in seconds the agent should collect and publish disk metrics.	The value of the global `metrics_collection_interval` .	A string specifying the number of seconds. Accepts only whole numbers.	This value overrides the global `metrics_collection_interval` property only for disk metrics.

Property	Required	Description	Default value	Possible values	Notes
`drop_original_metrics`	Optional	List of disk metrics for which to not publish unaggregated metrics.	No unaggregated disk metrics published.	A comma-separated list of disk metrics that are also specified in the metrics property. An empty string means to publish all disk metrics.	The CloudWatch agent aggregates all metrics by cluster ID, instance ID, node type, and service name. By default, the CloudWatch agent doesn't publish the per-resource metrics for metrics with multiple resources.

Property	Required	Description	Default value	Possible values	Notes
`resources`	Optional	Determines whether the agent will publish per-mount-point metrics.	"*"	"*" means all mount points, "" means no mount points, or a comma-separated list of mount points. For example, `"/,/emr"`.	The CloudWatch agent only publishes per-mount-point metrics for disk metrics that aren't dropped in `drop_original_metrics` .

diskio

Property	Required	Description	Default value	Possible values	Notes
`metrics`	Optional	The list of disk IO metrics for the agent to collect.	See [Default metrics for CloudWatch agent with Amazon EMR](#)	A comma-separated list of valid disk IO metric names with or without the `diskio_` prefix, such as `diskio_reads` and `writes`.	Specifying an empty string means to not publish any disk IO metrics.

Property	Required	Description	Default value	Possible values	Notes
				See [Metrics collected by the CloudWatch agent] for valid metrics.	
`metrics_collection_interval`	Optional	How often in seconds the agent should collect and publish disk IO metrics.	The value of the global `metrics_collection_interval` .	A string specifying the number of seconds. Accepts only whole numbers.	This value overrides the global `metrics_collection_interval` property only for disk IO metrics.

Property	Required	Description	Default value	Possible values	Notes
`drop_original_metrics`	Optional	List of disk IO metrics for which to not publish unaggregated metrics.	No unaggregated disk IO metrics published.	A comma-separated list of disk IO metrics that are also specified in the metrics property. An empty string means to publish all disk IO metrics.	The CloudWatch agent aggregates all metrics by cluster ID, instance ID, node type, and service name. By default, the CloudWatch agent doesn't publish the per-resource metrics for metrics with multiple resources.

Property	Required	Description	Default value	Possible values	Notes
`resources`	Optional	Determines whether the agent will publish per-device metrics.	"*"	"*" means all storage devices, "" means no storage devices, or a comma-separated list of device names. For example, `"nvme0n1, nvme1n1"`.	The CloudWatch agent only publishes per-device metrics for disk IO metrics that aren't dropped in `drop_original_metrics`.

mem

Property	Required	Description	Default value	Possible values	Notes
`metrics`	Optional	The list of memory metrics for the agent to collect.	See Default metrics for CloudWatch agent with Amazon EMR	A comma-separated list of valid memory metric names with or without the `mem_` prefix, such as `mem_available` and `available_percent`.	Specifying an empty string means to not publish any memory metrics.

Property	Required	Description	Default value	Possible values	Notes
				See Metrics collected by the CloudWatch agent for valid metrics.	
`metrics_collection_interval`	Optional	How often in seconds the agent should collect and publish memory metrics.	The value of the global `metrics_collection_interval`.	A string specifying the number of seconds. Accepts only whole numbers.	This value overrides the global `metrics_collection_interval` property only for memory metrics.

net

Property	Required	Description	Default value	Possible values	Notes
`metrics`	Optional	The list of network metrics for the agent to collect.	See Default metrics for CloudWatch agent with Amazon EMR	A comma-separated list of valid network metric names with or without the `net_` prefix,	Specifying an empty string means to not publish any network metrics.

Property	Required	Description	Default value	Possible values	Notes
				such as `net_packets_sent` and `packets_recv` . See [Metrics collected by the CloudWatch agent](#) for valid metrics.	
`metrics_collection_interval`	Optional	How often in seconds the agent should collect and publish network metrics.	The value of the global `metrics_collection_interval` .	A string specifying the number of seconds. Accepts only whole numbers.	This value overrides the global `metrics_collection_interval` property only for network metrics.

Property	Required	Description	Default value	Possible values	Notes
drop_original_metrics	Optional	List of network metrics for which to not publish unaggregated metrics.	No unaggregated network metrics published.	A comma-separated list of network metrics that are also specified in the metrics property. An empty string means to publish all network metrics.	The CloudWatch agent aggregates all metrics by cluster ID, instance ID, node type, and service name. By default, the CloudWatch agent doesn't publish the per-resource metrics for metrics with multiple resources.

Property	Required	Description	Default value	Possible values	Notes
`resources`	Optional	Determines whether the agent will publish per-interface metrics.	"*"	"*" means all network interfaces, "" means no network interfaces, or a comma-separated list of interfaces names. For example, `"eth0,eth1"`.	The CloudWatch agent only publishes per-interface metrics for network metrics that aren't dropped in `drop_original_metrics`.

netstat

Property	Required	Description	Default value	Possible values	Notes
`metrics`	Optional	The list of network statistics metrics for the agent to collect.	See [Default metrics for CloudWatch agent with Amazon EMR]()	A comma-separated list of valid memory metric names with or without the `netstat_` prefix, such as `tcp_listen` and	Specifying an empty string means to not publish any network statistic metrics.

Property	Required	Description	Default value	Possible values	Notes
				netstat_udp_socket . See [Metrics collected by the CloudWatch agent](#) for valid metrics.	
metrics_collection_interval	Optional	How often in seconds the agent should collect and publish network statistic metrics.	The value of the global metrics_collection_interval .	A string specifying the number of seconds. Accepts only whole numbers.	This value overrides the global metrics_collection_interval property only for network statistic metrics.

processes

Property	Required	Description	Default value	Possible values	Notes
metrics	Optional	The list of process metrics for the agent to collect.	See [Default metrics for CloudWatch agent with](#)	A comma-separated list of valid memory metric	Specifying an empty string means to not publish any

Property	Required	Description	Default value	Possible values	Notes
			Amazon EMR	names with or without the processes_ prefix, such as processes_running and total. See Metrics collected by the CloudWatch agent for valid metrics.	process metrics.
metrics_collection_interval	Optional	How often in seconds the agent should collect and publish system process metrics.	The value of the global metrics_collection_interval .	A string specifying the number of seconds. Accepts only whole numbers.	This value overrides the global metrics_collection_interval property only for system process metrics.

swap

Property	Required	Description	Default value	Possible values	Notes
`metrics`	Optional	The list of swap metrics for the agent to collect.	See [Default metrics for CloudWatch agent with Amazon EMR](#)	A comma-separated list of valid memory metric names with or without the `swap_` prefix, such as `swap_free` and `used_percent` . See [Metrics collected by the CloudWatch agent](#) for valid metrics.	Specifying an empty string means to not publish any swap metrics.
`metrics_collection_interval`	Optional	How often in seconds the agent should collect and publish swap metrics.	The value of the global `metrics_collection_interval` .	A string specifying the number of seconds. Accepts only whole numbers.	This value overrides the global `metrics_collection_interval` property only for

Property	Required	Description	Default value	Possible values	Notes
					swap metrics.

emr-hadoop-hdfs-datanode-metrics properties

Property	Required	Description	Default value	Possible values	
`<custom_bean_name>`	Optional	N/A	The MBean that CloudWatch agent should collect metrics from, such as `Hadoop:service=DataNode,name=DataNodeActivity` . You can find sample MBean names and their corresponding metrics in the [example JMX YAML files](#) for Amazon EMR release 7.0.	A string containing the comma-delimited list of metrics that are associated with the MBean. For example, `BlocksCached,BlocksRead` .	

Property	Required	Description	Default value	Possible values
`otel.metric.export.interval`	Optional	How often in milliseconds to collect Hadoop DataNode metrics.	"60000"	A string specifying the number of milliseconds. Accepts only whole numbers.

emr-hadoop-hdfs-namenode-metrics properties

Property	Required	Description	Default value	Possible values
<custom_bean_name>	Optional	N/A	The MBean that CloudWatch agent should collect metrics from, such as `Hadoop:service=NameNode,name=FSNamesystem` . You can find sample MBean names and their corresponding metrics in the	A string containing the comma-delimited list of metrics that are associated with the MBean. For example, `BlockCapacity,CapacityUsedGB` .

Property	Required	Description	Default value	Possible values
			example JMX YAML files for Amazon EMR release 7.0.	
`otel.metric.export.interval`	Optional	How often in milliseconds to collect Hadoop NameNode metrics.	"60000"	A string specifying the number of milliseconds. Accepts only whole numbers.

emr-hadoop-yarn-nodemanager-metrics properties

Property	Required	Description	Default value	Possible values
<custom_bean_name>	Optional	N/A	The MBean that CloudWatch agent should collect metrics from, such as `Hadoop:service=NodeManager,name=NodeManagerMetrics` . You can	A string containing the comma-delimited list of metrics that are associated with the MBean. For example, `MaxCapacity,AllocatedGB` .

Property	Required	Description	Default value	Possible values
			find sample MBean names and their correspon ding metrics in the [example JMX YAML files](#) for Amazon EMR release 7.0.	
`otel.metr ic.export .interval`	Optional	How often in milliseco nds to collect Hadoop YARN NodeManag er metrics.	"60000"	A string specifying the number of milliseco nds. Accepts only whole numbers.

emr-hadoop-yarn-resourcemanager-metrics properties

Property	Required	Description	Default value	Possible values
<custom_b ean_name>	Optional	N/A	The MBean that CloudWatch agent should collect metrics from, such as `Hadoop:se`	A string containing the comma-delimited list of metrics that are associate d with the

Property	Required	Description	Default value	Possible values
			`rvice=ResourceManager,name=PartitionQueueMetrics` . You can find sample MBean names and their corresponding metrics in the [example JMX YAML files](#) for Amazon EMR release 7.0.	MBean. For example, `MaxCapacity,MaxCapacityVCores` .
`otel.metric.export.interval`	Optional	How often in milliseconds to collect Hadoop YARN ResourceManager metrics.	"60000"	A string specifying the number of milliseconds. Accepts only whole numbers.

emr-hbase-master-metrics properties

Property	Required	Description	Default value	Possible values
<custom_bean_name>	Optional	N/A	The MBean that CloudWatch agent should collect metrics from, such as `Hadoop:service=HBase,name=Master,sub=AssignmentManager` . You can find sample MBean names and their corresponding metrics in the example JMX YAML files for Amazon EMR release 7.0.	A string containing the comma-delimited list of metrics that are associated with the MBean. For example, `AssignFailedCount`, `AssignSubmittedCount` .
`otel.metric.export.interval`	Optional	How often in milliseconds to collect HBase Master metrics.	"60000"	A string specifying the number of milliseconds. Accepts

Property	Required	Description	Default value	Possible values
				only whole numbers.

emr-hbase-region-server-metrics properties

Property	Required	Description	Default value	Possible values
`<custom_bean_name>`	Optional	N/A	The MBean that CloudWatch agent should collect metrics from, such as `Hadoop:service=HBase,name=RegionServer,sub=IPC` . You can find sample MBean names and their corresponding metrics in the [example JMX YAML files](#) for Amazon EMR release 7.0.	A string containing the comma-delimited list of metrics that are associated with the MBean. For example, `numActiveHandler,numActivePriorityHandler` .

Property	Required	Description	Default value	Possible values
otel.metric.export.interval	Optional	How often in milliseconds to collect HBase Region Server metrics.	"60000"	A string specifying the number of milliseconds. Accepts only whole numbers.

emr-hbase-rest-server-metrics properties

Property	Required	Description	Default value	Possible values
<custom_bean_name>	Optional	N/A	The MBean that CloudWatch agent should collect metrics from, such as `Hadoop:service=HBase,name=REST` . You can find sample MBean names and their corresponding metrics in the example JMX	A string containing the comma-delimited list of metrics that are associated with the MBean. For example, `successfulPut,successfulScanCount` .

Property	Required	Description	Default value	Possible values
			[YAML files](#) for Amazon EMR release 7.0.	
`otel.metric.export.interval`	Optional	How often in milliseconds to collect HBase Rest Server metrics.	"60000"	A string specifying the number of milliseconds. Accepts only whole numbers.

emr-hbase-thrift-server-metrics properties

Property	Required	Description	Default value	Possible values
<custom_bean_name>	Optional	N/A	The MBean that CloudWatch agent should collect metrics from, such as `Hadoop:service=HBase,name=Thrift,sub=ThriftOne` . You can find sample MBean	A string containing the comma-delimited list of metrics that are associated with the MBean. For example, `BatchGet_max,BatchGet_mean` .

Property	Required	Description	Default value	Possible values
			names and their correspon ding metrics in the [example JMX YAML files](#) for Amazon EMR release 7.0.	
`otel.metric.export.interval`	Optional	How often in milliseco nds to collect HBase Thrift server metrics.	"60000"	A string specifying the number of milliseco nds. Accepts only whole numbers.

System metrics configurations examples

The following example demonstrates how to configure the CloudWatch agent to stop exporting all system metrics.

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "emr-system-metrics",
        "Properties": {},
        "Configurations": []
      }
    ]
  }
```

```
]
```

The following example configures the CloudWatch agent to export the default system metrics. Doing so is a quick way to reset the agent back to only exporting the default system metrics if you've already reconfigured the system metrics at least once. This reset also removes any application metrics that were reconfigured before.

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": []
  }
]
```

The following example configures the cluster to export the `cpu`, `mem`, and the `disk` metrics.

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "emr-system-metrics",
        "Properties": {
          "metrics_collection_interval": "20"
        },
        "Configurations": [
          {
            "Classification": "cpu",
            "Properties": {
              "metrics": "cpu_usage_guest,cpu_usage_idle",
              "metrics_collection_interval": "30",
              "drop_original_metrics": "cpu_usage_guest"
            }
          },
          {
            "Classification": "mem",
            "Properties": {
              "metrics": "mem_active"
            }
          },
```

```
            {
             "Classification": "disk",
             "Properties": {
               "metrics": "disk_used_percent",
               "resources": "/,/mnt",
               "drop_original_metrics": ""
             }
           }
         ]
       }
     ]
   }
]
```

The previous example configuration has the following properties:

- Every 30 seconds, the agent collects the `cpu_guest` metric for all CPUs. You can find the aggregated metric under the CloudWatch namespace `CWAgent > cluster.id, instance.id, node.type, service.name`.
- Every 30 seconds, the agent collects the `cpu_idle` metric for all CPUs. You can find the aggregated metric under the CloudWatch namespace `CWAgent > cluster.id, instance.id, node.type, service.name`. The agent also collects the per-cpu metrics. You can find them in the same namespace. The agent collects this metric because the `drop_original_metrics` property doesn't contain `cpu_idle`, so the agent doesn't ignore the metric.
- Every 20 seconds, the agent collects the `mem_active` metric. You can find the aggregated metric under the CloudWatch namespace `CWAgent > cluster.id, instance.id, node.type, service.name`.
- Every 20 seconds, the agent collects the `disk_used_percent` metrics for the / and `/mnt` disk mounts. You can find the aggregated metrics under the CloudWatch namespace `CWAgent > cluster.id, instance.id, node.type, service.name`. The agent also collects the per-mount metrics. You can find them in the same namespace. The agent collects this metric because the `drop_original_metrics` property doesn't contain `disk_used_percent`, so the agent doesn't ignore the metric.

Application metrics configurations examples

The following example configures the CloudWatch agent to stop exporting metrics for the Hadoop Namenode service.

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "emr-hadoop-hdfs-namenode-metrics",
        "Properties": {},
        "Configurations": []
      }
    ]
  }
]
```

The following example configures a cluster to export Hadoop application metrics.

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "emr-hadoop-hdfs-namenode-metrics",
        "Properties": {
          "Hadoop:service=NameNode,name=FSNamesystem": "BlockCapacity,CapacityUsedGB",
          "otel.metric.export.interval": "20000"
        },
        "Configurations": []
      },
       {
        "Classification": "emr-hadoop-hdfs-datanode-metrics",
        "Properties": {
          "Hadoop:service=DataNode,name=JvmMetrics": "MemNonHeapUsedM",
          "otel.metric.export.interval": "30000"
        },
        "Configurations": []
      },
       {
```

```
        "Classification": "emr-hadoop-yarn-resourcemanager-metrics",
        "Properties": {
          "Hadoop:service=ResourceManager,name=CapacitySchedulerMetrics":
 "AllocateNumOps,NodeUpdateNumOps"
        },
        "Configurations": []
      }
    ]
  }
]
```

The previous example has the following properties:

- Every 20 seconds, the agent collects the `BlockCapacity` and `CapacityUsedGB` metrics from instances running the Hadoop Namenode service.
- Every 30 seconds, the agent collects `MemNonHeapUsedM` metrics from instances running the Hadoop Datanode service.
- Every 30 seconds, the agent collects the `AllocateNumOps` and `NodeUpdateNumOps` metrics from instances that run the Hadoop YARN ResourceManaager.

Amazon Managed Service for Prometheus example

The following example demonstrates how to configure the CloudWatch agent to export metrics to Amazon Managed Service for Prometheus.

If you are currently exporting metrics to Amazon Managed Service for Prometheus and want to reconfigure the metrics for the cluster and continue exporting metrics to Amazon Managed Service for Prometheus, you must include the properties `metrics_destination` and `prometheus_endpoint`.

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {
      "metrics_destination": "prometheus",
      "prometheus_endpoint": "http://amp-workspace/api/v1/remote_write"
    },
    "Configurations": []
  }
]
```

To use the CloudWatch agent to export metrics to CloudWatch, use the following example.

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {
      "metrics_destination": "cloudwatch"
    },
    "Configurations": []
  }
]
```

Note

The CloudWatch agent has a Prometheus exporter that renames certain attributes. For the default metrics labels, Amazon Managed Service for Prometheus uses underscore characters in place of the periods that Amazon CloudWatch uses. If you use Amazon Managed Grafana to visualize the default metrics in Amazon Managed Service for Prometheus, the labels appear as `cluster_id`, `instance_id`, `node_type`, and `service_name`.

Configure CloudWatch agent for Amazon EMR 7.0.0

You can configure the Amazon CloudWatch agent to use additional system metrics beyond those that the default CloudWatch agent configuration provides. The configuration for 7.0.0 requires the use of bootstrap actions, which we've provided examples for in the following sections. In an upcoming release, Amazon EMR will provide additional configuration options through the Amazon EMR API.

Topics

Configure additional system metrics with Amazon EMR 7.0.0

Use the following steps to configure the agent to use a different set of system metrics in Amazon EMR 7.0.0:

1. Create or choose a bucket in your Amazon S3 account where you want to store the configuration files that specify the CloudWatch agent metrics.
2. Create the `emr-amazon-cloudwatch-agent.json` configuration file with your preferred metrics specified. To do this, use one of the methods explained in Create the CloudWatch agent configuration file. For more information about the structure of the CloudWatch agent configuration file, see Manually create or edit the CloudWatch agent configuration file in the *Amazon CloudWatch User Guide*.
3. Next, navigate to the **aws-emr-utilities** repo on GitHub and download the following system metrics scripts:
    - **`install_system_metrics_launcher.sh`** – A script that downloads and then runs `install_system_metrics.sh` in the background so that the node can finish bootstrapping.
    - **`install_system_metrics.sh`** – A script that waits for the instance it runs on to finish bootstrapping, then downloads and applies the configuration in the JSON file.
4. Open each SH file and replace *`my-s3-bucket`* with the name of your bucket from Step 1.
5. Upload the one JSON and two SH files to your S3 bucket.
6. Now, you can navigate to the Amazon EMR console and create a new cluster with the CloudWatch agent. Under **EMR on EC2** in the left navigation, select **Clusters** and then **Create cluster**.
7. In the **Name and applications** section, choose an Amazon EMR release of 7.0.0 or higher.
8. Under **Application bundle**, select the bundle or apps that you want to install to your cluster, and include **Amazon CloudWatch Agent** with your selections.
9. In the **Bootstrap actions** section, select **Add**.
    - For the **Name**, insert **`install_system_metrics_launcher.sh`**.
    - For the **Script location**, insert **`s3://`*`my-s3-bucket`*`/install_system_metrics_launcher.sh`**. Replace *`my-s3-bucket`* with the path to your S3 bucket.
    - Leave the **Arguments** block empty.

10. Select **Add bootstrap action**.
11. Continue to create the cluster to serve your workload needs.

When your cluster launches, the CloudWatch agent publishes the system metrics that you specified in the configuration file to CloudWatch.

Configure application metrics with Amazon EMR 7.0.0

You can configure the Amazon CloudWatch agent to publish application metrics for HDFS and YARN in addition to system metrics. Use the following steps to configure the agent to publish application metrics:

1. Create or choose a bucket in your Amazon S3 account where you want to store the configuration files that specify the CloudWatch agent metrics.
2. Next, navigate to the **aws-emr-utilities** repo on GitHub and download the following scripts:
   - **`install_app_metrics_launcher.sh`** – A script that downloads and then runs `install_app_metrics.sh` in the background so that the node can finish bootstrapping.
   - **`install_app_metrics.sh`** – A script that waits for the instance it runs on to finish bootstrapping, then downloads and applies the configuration in the YAML files that you'll download in an upcoming step.
3. Open each file and replace *`my-s3-bucket`* with the name of your bucket from Step 1.
4. Next, download the following YAML mapping files. For information about how these YAML files are structured, see `javaagent` in the *OpenTelemetry Instrumentation for Java* GitHub repo.
   - **`datanode-metrics.yaml`** – The configuration for Hadoop DataNode metrics.
   - **`namenode-metrics.yaml`** – The configuration for Hadoop NameNode metrics.
   - **`nodemanager-metrics.yaml`** – The configuration for Yarn NodeManager metrics.
   - **`resourcemanager-metrics.yaml`** – The configuration for Yarn ResourceManager metrics.
5. Upload the two SH and four YAML files to your S3 bucket.
6. Now, you can navigate to the Amazon EMR console and create a new cluster with the CloudWatch agent. Under **EMR on EC2** in the left navigation, select **Clusters** and then **Create cluster**.
7. In the **Name and applications** section, choose an Amazon EMR release of 7.0.0 or higher.

8. Under **Application bundle**, select the bundle or custom group of apps that you want to install to your cluster, and include **CloudWatch agent** with your selections.
9. In the **Bootstrap actions** section, select **Add**.
   - For the **Name**, insert **`install_app_metrics_launcher.sh`**.
   - For the **Script location**, insert **`s3://my-s3-bucket/install_app_metrics_launcher.sh`**. Replace *`my-s3-bucket`* with the path to your S3 bucket.
   - Leave the **Arguments** block empty.
10. Select **Add bootstrap action**.
11. Continue to create the cluster to serve your workload needs.

When your cluster launches, The CloudWatch agent publishes the application metrics that you specified along with the system metrics to CloudWatch.

Configure Amazon Managed Service for Prometheus as cloud storage for metrics with Amazon EMR 7.0.0

You can configure the Amazon CloudWatch agent to publish metrics to Amazon Managed Service for Prometheus instead of CloudWatch.

> **Note**
>
> You can publish metrics from the Amazon CloudWatch agent to either Amazon Managed Service for Prometheus or to Amazon CloudWatch, but you can't publish the metrics to both services for the same cluster.

To configure the agent to publish metrics to Amazon Managed Service for Prometheus, you must add the `aps:RemoteWrite` AWS Identity and Access Management (IAM) permission to the Amazon EC2 instance profile for Amazon EMR. The following example policy contains the required permission:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
```

```
        "Effect": "Allow",
        "Action": "aps:RemoteWrite",
        "Resource": "*"
      }
    ]
}
```

Use the CloudWatch agent on an EMR cluster to publish metrics to Amazon Managed Service for Prometheus

Once the service policy has the correct permissions, use the following steps to launch a cluster that uses the CloudWatch agent to publish metrics to Amazon Managed Service for Prometheus.

1. Use the AWS Management Console or AWS CLI to create an Amazon Managed Service for Prometheus workspace. For more information, see Create a workspace in the *Amazon Managed Service for Prometheus User Guide*.
2. Create or choose a bucket in your Amazon S3 account where you want to store the launch files that specify Amazon Managed Service for Prometheus as cloud storage.
3. Next, navigate to the **aws-emr-utilities** repo on GitHub and download the following scripts:
   - **add_prometheus_endpoint_launcher.sh** – A script that downloads and then runs `add_prometheus_endpoint.sh` in the background so that the node can finish bootstrapping.
   - **add_prometheus_endpoint.sh** – A script that waits for the instance it runs on to finish bootstrapping, then configures CloudWatch agent to publish to the Amazon Managed Service for Prometheus endpoint that you provide as an argument when you launch your cluster.
4. Open each file and replace *my-s3-bucket* with the name of your bucket from Step 2.
5. Use the AWS CLI to create an EMR cluster with the `add_prometheus_endpoint_launcher.sh` bootstrap action. In the following command, replace *my-s3-bucket* with the bucket that holds the bootstrap action, and replace *managedpro-remote-write-workspace-url* with the remote write endpoint for your Amazon Managed Service for Prometheus workspace. Be sure to specify an Amazon EMR release label of `emr-7.0.0` or higher.

```
aws emr create-cluster --name managedpro-cluster \
    --release-label emr-7.0.0 \
    --applications Name=Hadoop Name=AmazonCloudWatchAgent \
```

```
    --ec2-attributes KeyName=myKey --instance-type m7g.2xlarge \
    --instance-count 3 --use-default-roles
   --bootstrap-actions Name='Add Prometheus Endpoint',Path=s3://my-s3-bucket/
add_prometheus_endpoint_launcher.sh,Args='managedpro-remote-write-workspace-url'
```

When your cluster launches, the CloudWatch agent publishes the metrics it collects to Amazon Managed Service for Prometheus.

Use Amazon Managed Service for Prometheus as a data source for Amazon Managed Grafana

Once Amazon EMR has published the cluster metrics to Amazon Managed Service for Prometheus, you can use the following steps to visualize the metrics with Amazon Managed Grafana:

1. Use the AWS Management Console to create an Amazon Managed Grafana workspace and user with appropriate permissions. For more information, see Create a workspace in the *Amazon Managed Grafana User Guide*.
2. Add your Amazon Managed Service for Prometheus workspace as a data source to Amazon Managed Grafana. For more information, see Use AWS data source configuration to add Amazon Managed Service for Prometheus as a data source in the *Amazon Managed Grafana User Guide*.

Note

The CloudWatch agent has a Prometheus exporter that renames certain attributes. For the default metrics labels, Amazon Managed Service for Prometheus uses underscore characters in place of the periods that Amazon CloudWatch uses. So if you use Amazon Managed Grafana to visualize the default metrics in Amazon Managed Service for Prometheus, the labels appear as `jobflow_id`, `instance_id`, and `service_name`. Also, any **application** metrics that the CloudWatch agent publishes to Amazon Managed Service for Prometheus use the label `job` instead of `service_name`. However, **system** metrics continue to use the `service_name` label.

Considerations and limitations

- The native Amazon CloudWatch agent is available for clusters that you create with Amazon EMR releases 7.0.0 and higher.

- The configuration for CloudWatch agent in Amazon EMR 7.0.0 requires the use of bootstrap actions. In an upcoming release, Amazon EMR will provide additional configuration options through the Amazon EMR API.
- You can't install the Amazon EMR CloudWatch agent if you've already deployed the CloudWatch agent by another method such as AWS Systems Manager Agent (SSM Agent) in the Region where you create your cluster. Doing so will result in your cluster terminating with errors.

CloudWatch agent release history

The following table lists the version of AmazonCloudWatchAgent included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in Amazon EMR 7.x release versions, Amazon EMR 6.x release versions, or Amazon EMR 5.x release versions.

AmazonCloudWatchAgent version information

Amazon EMR Release label	AmazonCloudWatchAgent Version	Components installed with AmazonCloudWatchAgent
emr-7.2.0	1.300032.2	adot-java-agent, emrfs, emr-amazon-cloudwatch-agent , emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-7.1.0	1.300032.2	adot-java-agent, emrfs, emr-amazon-cloudwatch-agent , emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-

Amazon EMR Release label	AmazonCloudWatchAgent Version	Components installed with AmazonCloudWatchAgent
		cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-7.0.0	1.300031.1	adot-java-agent, emr-amazon-cloudwatch-agent

Delta Lake

Delta Lake is a storage layer framework for lakehouse architectures commonly built on Amazon S3. With Amazon EMR releases 6.9.0 and higher, you can use Apache Spark 3.x on Amazon EMR clusters with Delta Lake tables. For more information about lakehouses with Delta Lake, see https://delta.io/.

The following table lists the version of Delta included in the latest release of the Amazon EMR 7.x series, along with the components that Amazon EMR installs with Delta.

For the version of components installed with Delta in this release, see Release 7.2.0 Component Versions.

Delta version information for emr-7.2.0

Amazon EMR Release Label	Delta Version	Components Installed With Delta
emr-7.2.0	Delta 3.1.0	Not available.

The following table lists the version of Delta included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with Delta.

For the version of components installed with Delta in this release, see Release 6.15.0 Component Versions.

Delta version information for emr-6.15.0

Amazon EMR Release Label	Delta Version	Components Installed With Delta
emr-6.15.0	Delta 2.4.0	Not available.

Introduction to Delta Lake

Delta Lake is an open-source project that helps implement modern data lake architectures commonly built on Amazon S3. Delta Lake offers the following capabilities:

- Atomic, consistent, isolated, durable (ACID) transactions on Spark. Readers see a consistent view of the table during a Spark job.
- Scalable metadata handling with distributed processing by Spark.
- Combines streaming and batch uses cases with the same Delta table.
- Automatic schema enforcement to avoid bad records during data ingestion.
- Time travel with data versioning.
- Supports merge, update, and delete operations for complex use cases like change data capture (CDC), streaming upserts, and more.

Using a cluster with Delta Lake installed

Topics

Use a Delta Lake cluster with Flink

With Amazon EMR release 6.11 and higher, you can use Delta Lake with your Flink cluster. The following examples use the AWS CLI to work with Delta Lake on an Amazon EMR Flink cluster.

> **Note**
>
> Amazon EMR supports the Flink DataStream API when you use Delta Lake with a Flink cluster.

Create a Delta Lake cluster

1. Create a file, `delta_configurations.json`, with the following content:

```
[{"Classification":"delta-defaults",
    "Properties":{"delta.enabled":"true"}}]
```

2. Create a cluster with the following configuration. Replace the `example Amazon S3 bucket path` and the `subnet ID` with your own.

```
aws emr create-cluster
--release-label emr-6.11.0
--applications Name=Flink
--configurations file://delta_configurations.json
--region us-east-1  --name My_Spark_Delta_Cluster
--log-uri  s3://DOC-EXAMPLE-BUCKET/
--instance-type m5.xlarge
--instance-count 3
--service-role EMR_DefaultRole_V2
--ec2-attributes
 InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef0
```

Initialize a Flink yarn session

To initialize a Flink yarn session, run the following command:

```
flink-yarn-session -d
```

Build a Flink job with Delta Lake

The following examples show how to use sbt or Maven to build your Flink job with Delta Lake.

sbt

sbt is a build tool for Scala that you can use with little to no configuration when you have small projects.

```
libraryDependencies ++= Seq(
  "io.delta" %% "delta-flink" % deltaConnectorsVersion % "provided",
  "io.delta" %% "delta-standalone" % deltaConnectorsVersion % "provided",
  "org.apache.flink" %% "flink-clients" % flinkVersion % "provided",
  "org.apache.flink" %% "flink-parquet" % flinkVersion % "provided",
  "org.apache.hadoop" % "hadoop-client" % hadoopVersion % "provided",
  "org.apache.flink" % "flink-table-common" % flinkVersion % "provided",
  "org.apache.flink" %% "flink-table-runtime" % flinkVersion % "provided")
```

Maven

Maven is an open-source build automation tool from the Apache Software Foundation. With Maven, you can build, publish, and deploy a Flink job with Delta Lake on Amazon EMR.

```
<project>
<properties>
    <scala.main.version>2.12</scala.main.version>
    <delta-connectors-version>0.6.0</delta-connectors-version>
    <flink-version>1.16.1</flink-version>
    <hadoop-version>3.1.0</hadoop-version>
</properties>

<dependencies>
    <dependency>
        <groupId>io.delta</groupId>
        <artifactId>delta-flink</artifactId>
        <version>$delta-connectors-version</version>
        <scope>provided</scope>
    </dependency>
    <dependency>
        <groupId>io.delta</groupId>
        <artifactId>delta-standalone_$scala-main-version</artifactId>
        <version>$delta-connectors-version</version>
        <scope>provided</scope>
    </dependency>
    <dependency>
        <groupId>org.apache.flink</groupId>
        <artifactId>flink-clients</artifactId>
        <version>$flink-version</version>
        <scope>provided</scope>
    </dependency>
    <dependency>
        <groupId>org.apache.flink</groupId>
        <artifactId>flink-parquet</artifactId>
        <version>$flink-version</version>
        <scope>provided</scope>
    </dependency>
    <dependency>
        <groupId>org.apache.hadoop</groupId>
        <artifactId>hadoop-client</artifactId>
        <version>$hadoop-version</version>
        <scope>provided</scope>
    </dependency>
```

```
    <dependency>
        <groupId>org.apache.flink</groupId>
        <artifactId>flink-table-common</artifactId>
        <version>$flink-version</version>
        <scope>provided</scope>
    </dependency>
    <dependency>
        <groupId>org.apache.flink</groupId>
        <artifactId>flink-table-runtime</artifactId>
        <version>$flink-version</version>
        <scope>provided</scope>
    </dependency>
</dependencies>
```

Write to a Delta table with Flink Datastream API

Use the following example to create a DeltaSink to write to the table with a `deltaTablePath`:

```
public static DataStream<RowData> createDeltaSink(
        DataStream<RowData> stream,
        String deltaTablePath,
        RowType rowType) {
    Configuration configuration = new Configuration();
    DeltaSink<RowData> deltaSink = DeltaSink
            .forRowData(
                    new org.apache.flink.core.fs.Path(deltaTablePath),
                    configuration,
                    rowType)
            .build();
    stream.sinkTo(deltaSink);
    return stream;
}
```

Read from a Delta table with Flink Datastream API

Use the following example to create a bounded DeltaSource to read from the table with a `deltaTablePath`:

```
public static DataStream<RowData> createBoundedDeltaSourceAllColumns(
        StreamExecutionEnvironment env,
        String deltaTablePath) {
```

```
    Configuration configuration = new Configuration();
    DeltaSource<RowData> deltaSource = DeltaSource
            .forBoundedRowData(
                    new org.apache.flink.core.fs.Path(deltaTablePath),
                    configuration)
            .build();

    return env.fromSource(deltaSource, WatermarkStrategy.noWatermarks(), "delta-
source");
}
```

Sink creation with multi-cluster support for Delta Lake standalone

Use the following example to create a DeltaSink to write to table with a `deltaTablePath` and multi cluster support:

```
public DataStream<RowData> createDeltaSink(
        DataStream<RowData> stream,
        String deltaTablePath) {
    Configuration configuration = new Configuration();
    configuration.set("spark.delta.logStore.s3.impl",
 "io.delta.storage.S3DynamoDBLogStore");
    configuration.set("spark.io.delta.storage.S3DynamoDBLogStore.ddb.tableName",
 "delta_log");
    configuration.set("spark.io.delta.storage.S3DynamoDBLogStore.ddb.region", "us-
east-1");

    DeltaSink<RowData> deltaSink = DeltaSink
        .forRowData(
            new Path(deltaTablePath),
            configuration,
            rowType)
        .build();
    stream.sinkTo(deltaSink);
    return stream;
}
```

Run the Flink job

Use the following command to run your job:

```
flink run FlinkJob.jar
```

Use a Delta Lake cluster with Trino

With Amazon EMR releases 6.9.0 and higher, you can use Delta Lake with your Trino cluster.

In this tutorial, we will use the AWS CLI to work with Delta Lake on Amazon EMR Trino cluster.

Create a Delta Lake cluster

1. Create a file, `delta_configurations.json`, and set values for your chosen catalog. For example, if you want to use the Hive metastore as your catalog, your file should have the following content:

```
[{"Classification":"delta-defaults",
    "Properties":{"delta.enabled":"true"}},
    {"Classification":"trino-connector-delta",
    "Properties":{"hive.metastore.uri":"thrift://localhost:9083"}}]
```

   If you want to use the AWS Glue Catalog as your store, your file should have the following content:

```
[{"Classification":"delta-defaults",
    "Properties":{"delta.enabled":"true"}},
    {"Classification":"trino-connector-delta",
    "Properties":{"hive.metastore":"glue"}}]
```

2. Create a cluster with the following configuration, replacing the **`example Amazon S3 bucket path`** and the **`subnet ID`** with your own.

```
aws emr create-cluster
    --release-label emr-6.9.0
    --applications Name=Trino
    --configurations file://delta_configurations.json
    --region us-east-1  --name My_Spark_Delta_Cluster
    --log-uri  s3://DOC-EXAMPLE-BUCKET/
    --instance-type m5.xlarge
    --instance-count 2
    --service-role EMR_DefaultRole_V2
    --ec2-attributes
 InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef0
```

Initialize Trino session for Delta Lake

To initialize Trino session, run the following command

```
trino-cli --catalog delta
```

Write to a Delta Lake table

Create and write to your table with the following SQL commands:

```
SHOW SCHEMAS;

CREATE TABLE default.delta_table (id  int, data varchar, category varchar) WITH
( location =  's3://DOC-EXAMPLE-BUCKET/<prefix>');

INSERT INTO default.delta_table VALUES  (1,'a','c1'), (2,'b','c2'), (3,'c','c3');
```

Read from a Delta Lake table

Read from your table with the following SQL command:

```
SELECT * from default.delta_table;
```

Use a Delta Lake cluster with Spark

Starting with Amazon EMR version 6.9.0, you can use Delta Lake with your Spark cluster without the need for bootstrap actions. For Amazon EMR releases 6.8.0 and lower, you can use bootstrap actions to pre-install the necessary dependencies.

The following examples use the AWS CLI to work with Delta Lake on an Amazon EMR Spark cluster.

To use Delta Lake on Amazon EMR with the AWS Command Line Interface, first create a cluster. For information on how to specify the Delta Lake classification with AWS Command Line Interface, see Supply a configuration using the AWS Command Line Interface when you create a cluster or Supply a configuration with the Java SDK when you create a cluster.

1. Create a file, `configurations.json`, with the following content:

```
[{"Classification":"delta-defaults",  "Properties":{"delta.enabled":"true"} }]
```

2. Create a cluster with the following configuration, replacing the example Amazon S3 **bucket path** and the **subnet ID** with your own.

```
aws emr create-cluster
     --release-label  emr-6.9.0
     --applications Name=Spark
     --configurations file://delta_configurations.json
     --region us-east-1
     --name My_Spark_Delta_Cluster
     --log-uri  s3://DOC-EXAMPLE-BUCKET/
     --instance-type m5.xlarge
     --instance-count 2
     --service-role EMR_DefaultRole_V2
     --ec2-attributes
 InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef0
```

Alternatively, you can create an Amazon EMR cluster and Spark application with the following files as JAR dependencies in a Spark job:

```
/usr/share/aws/delta/lib/delta-core.jar,
/usr/share/aws/delta/lib/delta-storage.jar,
/usr/share/aws/delta/lib/delta-storage-s3-dynamodb.jar
```

Note

If you use Amazon EMR releases 6.9.0 or higher, use `/usr/share/aws/delta/lib/delta-spark.jar` instead of `/usr/share/aws/delta/lib/delta-core.jar`.

For more information, see Submitting Applications.

To include a jar dependency in the Spark job, you can add the following configuration properties to the Spark application:

```
--conf “spark.jars=/usr/share/aws/delta/lib/delta-core.jar,
     /usr/share/aws/delta/lib/delta-storage.jar,
```

```
    /usr/share/aws/delta/lib/delta-storage-s3-dynamodb.jar"
```

For more information about Spark job dependencies, see Dependency Management.

If you use Amazon EMR releases 6.9.0 or higher, add the `/usr/share/aws/delta/lib/delta-spark.jar` configuration instead.

```
--conf “spark.jars=/usr/share/aws/delta/lib/delta-spark.jar,
     /usr/share/aws/delta/lib/delta-storage.jar,
     /usr/share/aws/delta/lib/delta-storage-s3-dynamodb.jar"
```

Initialize a Spark session for Delta Lake

The following examples show how to launch the interactive Spark shell, use Spark submit, or use Amazon EMR Notebooks to work with Delta Lake on Amazon EMR.

spark-shell

1. Connect to the primary node using SSH. For more information, see Connect to the primary node using SSH in the *Amazon EMR Management Guide*.
2. Enter the following command to launch the Spark shell. To use the PySpark shell, replace `spark-shell` with `pyspark`.

```
spark-shell \
   --conf "spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension" \
   --conf
 "spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog"
```

If you run Amazon EMR releases 6.15.0 or higher, you must also use the following configurations to use fine-grained access control based on Lake Formation with Delta Lake.

```
spark-shell \
  --conf
 spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension,com.amazonaws.emr.recordserv
 \
  --conf
 spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog \

  --conf spark.sql.catalog.spark_catalog.lf.managed=true
```

spark-submit

1. Connect to the primary node using SSH. For more information, see [Connect to the primary node using SSH] in the *Amazon EMR Management Guide*.
2. Enter the following command to launch the Spark session for Delta Lake.

```
spark-submit
—conf "spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension"
—conf
 "spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog"
```

If you run Amazon EMR releases 6.15.0 or higher, you must also use the following configurations to use fine-grained access control based on Lake Formation with Delta Lake.

```
spark-submit \  `
--conf
 spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension,com.amazonaws.emr.recordserv
--conf
 spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog \

--conf spark.sql.catalog.spark_catalog.lf.managed=true
```

EMR Studio notebooks

To initialize a Spark session using Amazon EMR Studio notebooks, configure your Spark session using **%%configure** magic command in your Amazon EMR notebook, as in the following example. For more information, see [Use EMR Notebooks magics] in the *Amazon EMR Management Guide*.

```
%%configure -f
{
  "conf": {
    "spark.sql.extensions":  "io.delta.sql.DeltaSparkSessionExtension",
     "spark.sql.catalog.spark_catalog":
 "org.apache.spark.sql.delta.catalog.DeltaCatalog"
  }
}
```

If you run Amazon EMR releases 6.15.0 or higher, you must also use the following configurations to use fine-grained access control based on Lake Formation with Delta Lake.

```
%%configure -f
{
  "conf": {
    "spark.sql.extensions":
 "io.delta.sql.DeltaSparkSessionExtension,com.amazonaws.emr.recordserver.connector.spark.sql
    "spark.sql.catalog.spark_catalog":
 "org.apache.spark.sql.delta.catalog.DeltaCatalog",
    "spark.sql.catalog.spark_catalog.lf.managed": "true"
  }
}
```

Write to a Delta Lake table

The following example shows how to create a DataFrame and write it as a Delta Lake dataset. The example shows how to work with datasets with the Spark shell while connected to the primary node using SSH as the default hadoop user.

> **Note**
>
> To paste code samples into the Spark shell, type **:paste** at the prompt, paste the example, and then press **CTRL + D**.

PySpark

Spark includes a Python based shell, `pyspark`, that you can use to prototype Spark programs written in Python. Just as with `spark-shell`, invoke `pyspark` on the primary node.

```
## Create a DataFrame
data =  spark.createDataFrame([("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
("101",  "2015-01-01", "2015-01-01T12:14:58.597216Z"),
("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
("103",  "2015-01-01",  "2015-01-01T13:51:40.519832Z")],
["id", "creation_date",  "last_update_time"])

## Write a DataFrame as a Delta Lake dataset to the S3  location
spark.sql("""CREATE  TABLE IF NOT EXISTS delta_table (id string, creation_date
 string,
```

```
last_update_time string)
USING delta location
's3://DOC-EXAMPLE-BUCKET/example-prefix/db/delta_table'""");

data.writeTo("delta_table").append()
```

Scala

```
import org.apache.spark.sql.SaveMode
import org.apache.spark.sql.functions._

// Create a DataFrame
val data = Seq(("100",  "2015-01-01",  "2015-01-01T13:51:39.340396Z"),
("101",  "2015-01-01",  "2015-01-01T12:14:58.597216Z"),
("102",  "2015-01-01", "2015-01-01T13:51:40.417052Z"),
("103",  "2015-01-01",  "2015-01-01T13:51:40.519832Z")).toDF("id", "creation_date",
 "last_update_time")

// Write a DataFrame as a Delta Lake dataset to the S3  location
spark.sql("""CREATE  TABLE IF NOT EXISTS delta_table (id string,
creation_date string,
last_update_time string)
USING delta location
's3://DOC-EXAMPLE-BUCKET/example-prefix/db/delta_table'""");

data.write.format("delta").mode("append").saveAsTable("delta_table")
```

SQL

```
-- Create a Delta  Lake table with the S3 location
CREATE TABLE delta_table(id string,
creation_date string,
last_update_time string)
USING delta LOCATION
's3://DOC-EXAMPLE-BUCKET/example-prefix/db/delta_table';

-- insert data into the table
INSERT INTO delta_table VALUES  ("100", "2015-01-01",
 "2015-01-01T13:51:39.340396Z"),
("101",  "2015-01-01",  "2015-01-01T12:14:58.597216Z"),
("102",  "2015-01-01", "2015-01-01T13:51:40.417052Z"),
("103",  "2015-01-01", "2015-01-01T13:51:40.519832Z");
```

Read from a Delta Lake table

PySpark

```
ddf = spark.table("delta_table")
ddf.show()
```

Scala

```
val ddf =  spark.table("delta_table")
ddf.show()
```

SQL

```
SELECT * FROM delta_table;
```

Use a Delta Lake cluster with Spark and AWS Glue

To use the AWS Glue Catalog as the Metastore for Delta Lake tables, create a cluster with following steps. For information on specifying the Delta Lake classification using AWS Command Line Interface, see [Supply a configuration using the AWS Command Line Interface when you create a cluster] or [Supply a configuration using the Java SDK when you create a cluster].

Create a Delta Lake cluster

1. Create a file, `configurations.json`, with the following content:

```
[{"Classification":"delta-defaults",
"Properties":{"delta.enabled":"true"}},
{"Classification":"spark-hive-site",
"Properties":
{"hive.metastore.client.factory.class":"com.amazonaws.glue.catalog.metastore.AWSGlueDataCat
```

2. Create a cluster with the following configuration, replacing the **`example Amazon S3 bucket path`** and the **`subnet ID`** with your own.

```
aws emr create-cluster
```

```
    --release-label  emr-6.9.0
    --applications Name=Spark
    --configurations file://delta_configurations.json
    --region us-east-1
    --name My_Spark_Delta_Cluster
    --log-uri  s3://DOC-EXAMPLE-BUCKET/
    --instance-type m5.xlarge
    --instance-count 2
    --service-role EMR_DefaultRole_V2
    --ec2-attributes 
InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef0
```

Considerations and limitations

- Delta Lake is supported for use with Amazon EMR releases 6.9.0 and higher. You can use Apache Spark 3.x on Amazon EMR clusters with Delta tables.
- We recommend that you use s3 URI scheme for S3 location paths instead of s3a for best performance, security and reliability. For more information see Working with storage and file systems.
- With Amazon EMR 7.0, Delta Universal Format (UniForm) and convert-to-Iceberg statements aren't supported.
- With Amazon EMR 6.9 and 6.10, when you store Delta Lake table data in Amazon S3, column data becomes `NULL` after column rename operation. This issue is resolved with Amazon EMR 6.11. For more information about the experimental column rename operation, see Column rename operation in the Delta Lake User Guide.
- If you create a database in the AWS Glue Data Catalog outside of Apache Spark, the database could have an empty `LOCATION` field. Because Spark doesn't allow databases to be created with an empty location property, you'll get the following error if you use Spark in Amazon EMR to create a Delta table in a Glue database and the database has an empty `LOCATION` property:

  ```
  IllegalArgumentException: Can not create a Path from an empty string
  ```

  To resolve this issue, create the database in the Data Catalog with a valid, non-empty path for the `LOCATION` field. For steps to implement this solution, see Illegal argument exception when creating a table in the *Amazon Athena User Guide*.

Delta release history

The following table lists the version of Delta included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in Amazon EMR 7.x release versions, Amazon EMR 6.x release versions, or Amazon EMR 5.x release versions.

Delta version information

Amazon EMR Release label	Delta Version	Components installed with Delta
emr-7.2.0	3.1.0	Not available.
emr-7.1.0	3.0.0	Not available.
emr-7.0.0	3.0.0	Not available.
emr-6.15.0	2.4.0	Not available.
emr-6.14.0	2.4.0	Not available.
emr-6.13.0	2.4.0	Not available.
emr-6.12.0	2.4.0	Not available.
emr-6.11.1	2.2.0	Not available.
emr-6.11.0	2.2.0	Not available.
emr-6.10.1	2.2.0	Not available.
emr-6.10.0	2.2.0	Not available.
emr-6.9.1	2.1.0	Not available.
emr-6.9.0	2.1.0	Not available.

Apache Flink

Apache Flink is a streaming dataflow engine that you can use to run real-time stream processing on high-throughput data sources. Flink supports event time semantics for out-of-order events, exactly-once semantics, backpressure control, and APIs optimized to write both streaming and batch applications.

Additionally, Flink has connectors for third-party data sources, such as the following:

- Amazon Kinesis Data Streams
- Apache Kafka
- Flink Elasticsearch Connector
- Twitter Streaming API
- Cassandra

Amazon EMR supports Flink as a YARN application so that you can manage resources along with other applications within a cluster. Flink-on-YARN allows you to submit transient Flink jobs, or you can create a long-running cluster that accepts multiple jobs and allocates resources according to the overall YARN reservation.

Flink is included in Amazon EMR release versions 5.1.0 and later.

> **Note**
>
> Support for the `FlinkKinesisConsumer` class was added in Amazon EMR release version 5.2.1.

The following table lists the version of Flink included in the latest release of the Amazon EMR 7.x series, along with the components that Amazon EMR installs with Flink.

For the version of components installed with Flink in this release, see Release 7.2.0 Component Versions.

Flink version information for emr-7.2.0

Amazon EMR Release Label	Flink Version	Components Installed With Flink
emr-7.2.0	Flink 1.18.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta

The following table lists the version of Flink included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with Flink.

For the version of components installed with Flink in this release, see Release 6.15.0 Component Versions.

Flink version information for emr-6.15.0

Amazon EMR Release Label	Flink Version	Components Installed With Flink
emr-6.15.0	Flink 1.17.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client,

Amazon EMR Release Label	Flink Version	Components Installed With Flink
		flink-jobmanager-config, hudi, delta-standalone-connectors

The following table lists the version of Flink included in the latest release of the Amazon EMR 5.x series, along with the components that Amazon EMR installs with Flink.

For the version of components installed with Flink in this release, see Release 5.36.2 Component Versions.

Flink version information for emr-5.36.2

Amazon EMR Release Label	Flink Version	Components Installed With Flink
emr-5.36.2	Flink 1.14.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config

Topics

- Creating a cluster with Flink
- Configuring Flink in Amazon EMR
- Working with Flink jobs in Amazon EMR
- Use the Scala shell
- Finding the Flink web interface

- Flink autoscaler
- Optimizing job restart times for task recovery and scaling operations
- Working with Flink jobs from Zeppelin in Amazon EMR
- Flink release history

Creating a cluster with Flink

You can launch a cluster with the AWS Management Console, AWS CLI, or an AWS SDK.

To launch a cluster with Flink installed from the console

1. Open the Amazon EMR console at https://console.aws.amazon.com/emr.
2. Choose **Create cluster**, **Go to advanced options**.
3. For **Software Configuration**, choose **EMR Release emr-5.1.0** or later.
4. Choose **Flink** as an application, along with any others to install.
5. Select other options as necessary and choose **Create cluster**.

To launch a cluster with Flink from the AWS CLI

- Create the cluster with the following command:

```
aws emr create-cluster --release-label emr-7.2.0 \
--applications Name=Flink \
--configurations file://./configurations.json \
--region us-east-1 \
--log-uri s3://myLogUri \
--instance-type m5.xlarge \
--instance-count 2 \
--service-role EMR_DefaultRole_V2 \
--ec2-attributes KeyName=MyKeyName,InstanceProfile=EMR_EC2_DefaultRole \
--steps Type=CUSTOM_JAR,Jar=command-runner.jar,Name=Flink_Long_Running_Session,\
Args=flink-yarn-session,-d
```

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

Configuring Flink in Amazon EMR

Configure Flink with Hive Metastore and Glue Catalog

Amazon EMR releases 6.9.0 and higher support both Hive Metastore and AWS Glue Catalog with the Apache Flink connector to Hive. This section outlines the steps required to configure AWS Glue Catalog and Hive Metastore with Flink.

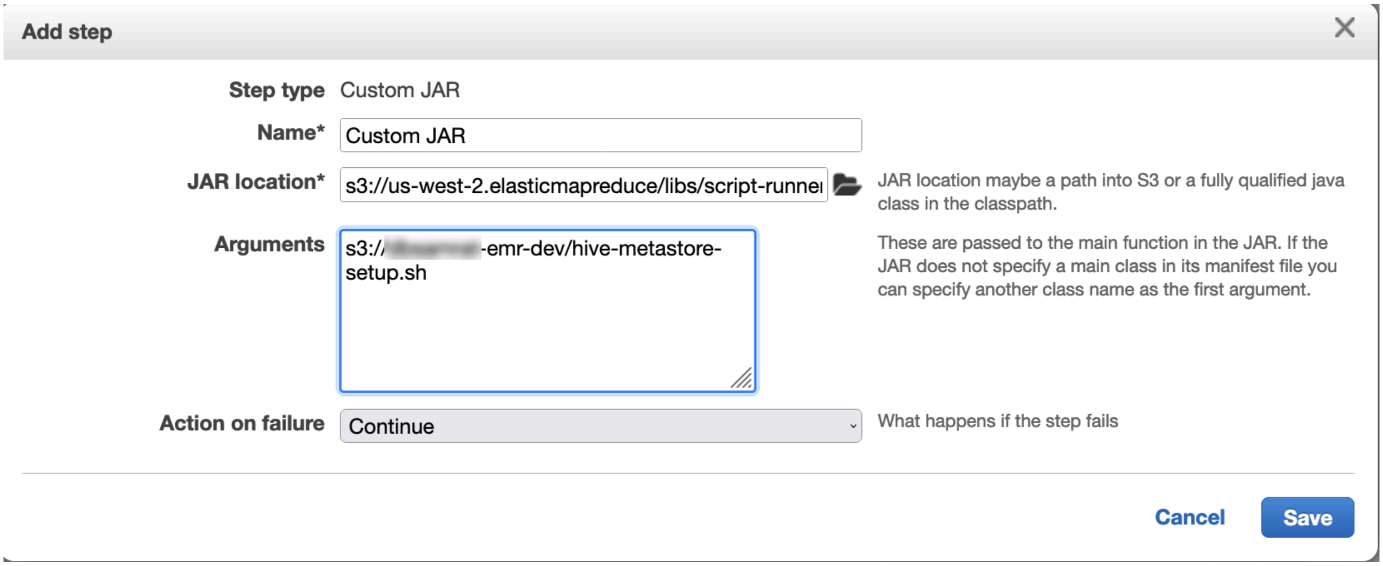
Topics

Use the Hive Metastore

1. Create an EMR cluster with release 6.9.0 or higher and at least two applications: **Hive** and **Flink**.
2. Use script runner to execute the following script as a step function:

   `hive-metastore-setup.sh`

```
sudo cp /usr/lib/hive/lib/antlr-runtime-3.5.2.jar /usr/lib/flink/lib
sudo cp /usr/lib/hive/lib/hive-exec-3.1.3*.jar /lib/flink/lib
sudo cp /usr/lib/hive/lib/libfb303-0.9.3.jar /lib/flink/lib
sudo cp /usr/lib/flink/opt/flink-connector-hive_2.12-1.15.2.jar /lib/flink/lib
sudo chmod 755 /usr/lib/flink/lib/antlr-runtime-3.5.2.jar
sudo chmod 755 /usr/lib/flink/lib/hive-exec-3.1.3*.jar
sudo chmod 755 /usr/lib/flink/lib/libfb303-0.9.3.jar
sudo chmod 755 /usr/lib/flink/lib/flink-connector-hive_2.12-1.15.2.jar
```

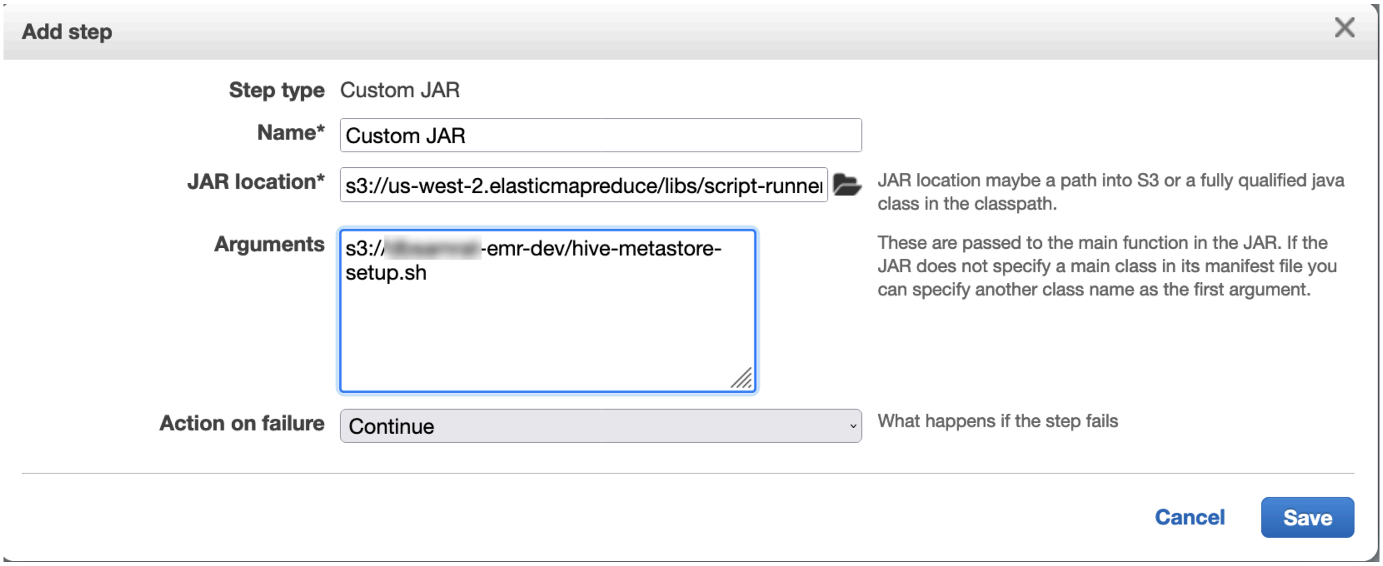


Use the AWS Glue Data Catalog

1. Create an EMR cluster with release 6.9.0 or higher and at least two applications: **Hive** and **Flink**.
2. Select **Use for Hive table metadata** in the AWS Glue Data Catalog settings to enable Data Catalog in the cluster.
3. Use script runner to execute the following script as a step function: Run commands and scripts on an Amazon EMR cluster:

   glue-catalog-setup.sh

```
sudo cp /usr/lib/hive/auxlib/aws-glue-datacatalog-hive3-client.jar /usr/lib/flink/
lib
sudo cp /usr/lib/hive/lib/antlr-runtime-3.5.2.jar /usr/lib/flink/lib
sudo cp /usr/lib/hive/lib/hive-exec-3.1.3*.jar /lib/flink/lib
sudo cp /usr/lib/hive/lib/libfb303-0.9.3.jar /lib/flink/lib
sudo cp /usr/lib/flink/opt/flink-connector-hive_2.12-1.15.2.jar /lib/flink/lib
sudo chmod 755 /usr/lib/flink/lib/aws-glue-datacatalog-hive3-client.jar
sudo chmod 755 /usr/lib/flink/lib/antlr-runtime-3.5.2.jar
sudo chmod 755 /usr/lib/flink/lib/hive-exec-3.1.3*.jar
sudo chmod 755 /usr/lib/flink/lib/libfb303-0.9.3.jar
sudo chmod 755 /usr/lib/flink/lib/flink-connector-hive_2.12-1.15.2.jar
```



Configure Flink with a configuration file

You can use the Amazon EMR configuration API to configure Flink with a configuration file. The files that are configurable within the API are:

- `flink-conf.yaml`
- `log4j.properties`
- `flink-log4j-session`
- `log4j-cli.properties`

The main configuration file for Flink is `flink-conf.yaml`.

To configure the number of task slots that are used for Flink from the AWS CLI

1. Create a file, `configurations.json`, with the following content:

```
[
    {
      "Classification": "flink-conf",
      "Properties": {
        "taskmanager.numberOfTaskSlots":"2"
      }
    }
```

```
]
```

2. Next, create a cluster with the following configuration:

```
aws emr create-cluster --release-label emr-7.2.0 \
--applications Name=Flink \
--configurations file://./configurations.json \
--region us-east-1 \
--log-uri s3://myLogUri \
--instance-type m5.xlarge \
--instance-count 2 \
--service-role EMR_DefaultRole_V2 \
--ec2-attributes KeyName=YourKeyName,InstanceProfile=EMR_EC2_DefaultRole
```

> **Note**
>
> You can also change some configurations with the Flink API. For more information, see *Concepts* in the Flink documentation.
> With Amazon EMR version 5.21.0 and later, you can override cluster configurations and specify additional configuration classifications for each instance group in a running cluster. You do this by using the Amazon EMR console, the AWS Command Line Interface (AWS CLI), or the AWS SDK. For more information, see Supplying a Configuration for an Instance Group in a Running Cluster.

Parallelism options

As the owner of your application, you know best what resources to assign to tasks within Flink. For the examples in this documentation, use the same number of tasks as the tasks instances that you use for the application. We generally recommend this for the initial level of parallelism, but you can also increase the granularity of parallelism with task slots, which should generally not exceed the number of virtual cores per instance. For more information about the Flink architecture, see *Concepts* in the Flink documentation.

Configuring Flink on an EMR cluster with multiple primary nodes

The JobManager of Flink remains available during the primary node failover process in an Amazon EMR cluster with multiple primary nodes. Beginning with Amazon EMR 5.28.0, JobManager high availability is also enabled automatically. No manual configuration is needed.

With Amazon EMR versions 5.27.0 or earlier, the JobManager is a single point of failure. When the JobManager fails, it loses all job states and will not resume the running jobs. You can enable JobManager high availability by configuring application attempt count, checkpointing, and enabling ZooKeeper as state storage for Flink, as the following example demonstrates:

```
[
  {
    "Classification": "yarn-site",
    "Properties": {
      "yarn.resourcemanager.am.max-attempts": "10"
    }
  },
  {
    "Classification": "flink-conf",
    "Properties": {
        "yarn.application-attempts": "10",
        "high-availability": "zookeeper",
        "high-availability.zookeeper.quorum": "%{hiera('hadoop::zk')}",
        "high-availability.storageDir": "hdfs:///user/flink/recovery",
        "high-availability.zookeeper.path.root": "/flink"
    }
  }
]
```

You must configure both maximum application master attempts for YARN and application attempts for Flink. For more information, see Configuration of YARN cluster high availability. You may also want to configure Flink checkpointing to make restarted JobManager recover running jobs from previously completed checkpoints. For more information, see Flink checkpointing.

Configuring memory process size

For Amazon EMR versions that use Flink 1.11.x, you must configure the total memory process size for both JobManager (`jobmanager.memory.process.size`) and TaskManager (`taskmanager.memory.process.size`) in `flink-conf.yaml`. You can set these values by either configuring the cluster with the configuration API or manually uncommenting these fields via SSH. Flink provides the following default values.

- `jobmanager.memory.process.size`: 1600m
- `taskmanager.memory.process.size`: 1728m

To exclude JVM metaspace and overhead, use the total Flink memory size (`taskmanager.memory.flink.size`) instead of `taskmanager.memory.process.size`. The default value for `taskmanager.memory.process.size` is 1280m. It's not recommended to set both `taskmanager.memory.process.size` and `taskmanager.memory.process.size`.

All Amazon EMR versions that use Flink 1.12.0 and later have the default values listed in the open-source set for Flink as the default values on Amazon EMR, so you don't need to configure them yourself.

Configuring log output file size

Flink application containers create and write to three types of log files: `.out` files, `.log` files, and `.err` files. Only `.err` files are compressed and removed from the file system, while `.log` and `.out` log files remain in the file system. To ensure these output files remain manageable and the cluster remains stable, you can configure log rotation in `log4j.properties` to set a maximum number of files and limit their sizes.

Amazon EMR versions 5.30.0 and later

Starting with Amazon EMR 5.30.0, Flink uses the log4j2 logging framework with the configuration classification name `flink-log4j`. The following example configuration demonstrates the log4j2 format.

```
[
  {
    "Classification": "flink-log4j",
    "Properties": {
      "appender.main.name": "MainAppender",
      "appender.main.type": "RollingFile",
      "appender.main.append" : "false",
      "appender.main.fileName" : "${sys:log.file}",
      "appender.main.filePattern" : "${sys:log.file}.%i",
      "appender.main.layout.type" : "PatternLayout",
      "appender.main.layout.pattern" : "%d{yyyy-MM-dd HH:mm:ss,SSS} %-5p %-60c %x - %m
%n",
      "appender.main.policies.type" : "Policies",
      "appender.main.policies.size.type" : "SizeBasedTriggeringPolicy",
      "appender.main.policies.size.size" : "100MB",
      "appender.main.strategy.type" : "DefaultRolloverStrategy",
      "appender.main.strategy.max" : "10"
    },
```

```
  }
]
```

Amazon EMR versions 5.29.0 and earlier

With Amazon EMR versions 5.29.0 and earlier, Flink uses the log4j logging framework. The following example configuration demonstrates the log4j format.

```
[
  {
    "Classification": "flink-log4j",
    "Properties": {
      "log4j.appender.file": "org.apache.log4j.RollingFileAppender",
      "log4j.appender.file.append":"true",
      # keep up to 4 files and each file size is limited to 100MB
      "log4j.appender.file.MaxFileSize":"100MB",
      "log4j.appender.file.MaxBackupIndex":4,
      "log4j.appender.file.layout":"org.apache.log4j.PatternLayout",
      "log4j.appender.file.layout.ConversionPattern":"%d{yyyy-MM-dd HH:mm:ss,SSS} %-5p
 %-60c %x - %m%n"
    },
  }
]
```

Configure Flink to run with Java 11

Amazon EMR releases 6.12.0 and higher provide Java 11 runtime support for Flink. The following sections describe how to configure the cluster to provide Java 11 runtime support for Flink.

Topics

Configure Flink for Java 11 when you create a cluster

Use the following steps to create an EMR cluster with Flink and Java 11 runtime. The configuration file where you add Java 11 runtime support is `flink-conf.yaml`.

Console

To create a cluster with Flink and Java 11 runtime in the console

1. Sign in to the AWS Management Console, and open the Amazon EMR console at https://console.aws.amazon.com/emr.
2. Choose **Clusters** under **EMR on EC2** in the navigation pane, and then **Create cluster**.
3. Select Amazon EMR release 6.12.0 or higher, and choose to install the Flink application. Select any other applications that you want to install on your cluster.
4. Continue setting up your cluster. In the optional **Software settings** section, use the default **Enter configuration** option and enter the following configuration:

```
[
    {
      "Classification": "flink-conf",
      "Properties": {
        "containerized.taskmanager.env.JAVA_HOME":"/usr/lib/jvm/jre-11",
        "containerized.master.env.JAVA_HOME":"/usr/lib/jvm/jre-11",
        "env.java.home":"/usr/lib/jvm/jre-11"
      }
    }
]
```

5. Continue to set up and launch your cluster.

AWS CLI

To create a cluster with Flink and Java 11 runtime from the CLI

1. Create a configuration file `configurations.json`that configures Flink to use Java 11.

```
[
    {
      "Classification": "flink-conf",
      "Properties": {
        "containerized.taskmanager.env.JAVA_HOME":"/usr/lib/jvm/jre-11",
        "containerized.master.env.JAVA_HOME":"/usr/lib/jvm/jre-11",
        "env.java.home":"/usr/lib/jvm/jre-11"
      }
    }
```

```
]
```

2. From the AWS CLI, create a new EMR cluster with Amazon EMR release 6.12.0 or higher, and install the Flink application, as shown in the following example:

```
aws emr create-cluster --release-label emr-6.12.0 \
--applications Name=Flink \
--configurations file://./configurations.json \
--region us-east-1 \
--log-uri s3://myLogUri \
--instance-type m5.xlarge \
--instance-count 2 \
--service-role EMR_DefaultRole_V2 \
--ec2-attributes KeyName=YourKeyName,InstanceProfile=EMR_EC2_DefaultRole
```

Configure Flink for Java 11 on a running cluster

Use the following steps to update a running EMR cluster with Flink and Java 11 runtime. The configuration file where you add Java 11 runtime support is `flink-conf.yaml`.

Console

To update a running cluster with Flink and Java 11 runtime in the console

1. Sign in to the AWS Management Console, and open the Amazon EMR console at https://console.aws.amazon.com/emr.
2. Choose **Clusters** under **EMR on EC2** in the navigation pane, and then select the cluster that you want to update.

> **Note**
>
> The cluster must use Amazon EMR release 6.12.0 or higher to support Java 11.

3. Select the **Configurations** tab.
4. In the **Instance group configurations** section, select the **Running** instance group that you want to update and then choose **Reconfigure** from the list actions menu.
5. Reconfigure the instance group with the **Edit attributes** option as follows. Select **Add new configuration** after each one.

Classification	Property	Value
flink-conf	containerized.taskmanager.env.JAVA_HOME	/usr/lib/jvm/jre-11
flink-conf	containerized.master.env.JAVA_HOME	/usr/lib/jvm/jre-11
flink-conf	env.java.home	/usr/lib/jvm/jre-11

6. Select **Save changes** to add the configurations.

AWS CLI

To update a running cluster to use Flink and Java 11 runtime from the CLI

Use the `modify-instance-groups` command to specify a new configuration for an instance group in a running cluster.

1. First, create a configuration file `configurations.json`that configures Flink to use Java 11. In the following example, replace *ig-1xxxxxxx9* with the ID for the instance group that you want to reconfigure. Save the file in the same directory where you will run the `modify-instance-groups` command.

```
[
   {
      "InstanceGroupId":"ig-1xxxxxxx9",
      "Configurations":[
         {
            "Classification":"flink-conf",
            "Properties":{
              "containerized.taskmanager.env.JAVA_HOME":"/usr/lib/jvm/jre-11",
              "containerized.master.env.JAVA_HOME":"/usr/lib/jvm/jre-11",
              "env.java.home":"/usr/lib/jvm/jre-11"
            },
            "Configurations":[]
         }
```

```
        ]
    }
]
```

2. From the AWS CLI, run the following command. Replace the ID for the instance group that you want to reconfigure:

```
aws emr modify-instance-groups --cluster-id j-2AL4XXXXXX5T9 \
--instance-groups file://configurations.json
```

Confirm the Java runtime for Flink on a running cluster

To determine the Java runtime for a running cluster, log in to the primary node with SSH as described in Connect to the primary node with SSH. Then run the following command:

```
ps -ef | grep flink
```

The `ps` command with the `-ef` option lists all running processes on the system. You can filter that output with `grep` to find mentions of the string `flink`. Review the output for the Java Runtime Environment (JRE) value, `jre-XX`. In the following output, `jre-11` indicates that Java 11 is picked up at runtime for Flink.

```
flink    19130     1  0 09:17 ?        00:00:15 /usr/lib/jvm/jre-11/bin/
java -Djava.io.tmpdir=/mnt/tmp -Dlog.file=/usr/lib/flink/log/flink-flink-
historyserver-0-ip-172-31-32-127.log -Dlog4j.configuration=file:/usr/lib/flink/conf/
log4j.properties -Dlog4j.configurationFile=file:/usr/lib/flink/conf/log4j.properties
 -Dlogback.configurationFile=file:/usr/lib/flink/conf/logback.xml -classpath /usr/lib/
flink/lib/flink-cep-1.17.0.jar:/usr/lib/flink/lib/flink-connector-files-1.17.0.jar:/
usr/lib/flink/lib/flink-csv-1.17.0.jar:/usr/lib/flink/lib/flink-json-1.17.0.jar:/usr/
lib/flink/lib/flink-scala_2.12-1.17.0.jar:/usr/lib/flink/lib/flink-table-api-java-
uber-1.17.0.jar:/usr/lib/flink/lib/flink-table-api-scala-bridge_2.12-1.17.0.
```

Alternatively, log in to the primary node with SSH and start a Flink YARN session with command `flink-yarn-session -d`. The output shows the Java Virtual Machine (JVM) for Flink, `java-11-amazon-corretto` in the following example:

```
2023-05-29 10:38:14,129 INFO  org.apache.flink.configuration.GlobalConfiguration
     [] - Loading configuration property: containerized.master.env.JAVA_HOME, /usr/lib/
jvm/java-11-amazon-corretto.x86_64
```

Working with Flink jobs in Amazon EMR

There are several ways to interact with Flink on Amazon EMR: through the console, the Flink interface found on the ResourceManager Tracking UI, and at the command line. You can submit a JAR file to a Flink application with any of these. Once submit a JAR file, it becomes a job that is managed by the Flink JobManager. The JobManager is located on the YARN node that hosts the Flink session Application Master daemon.

You can run a Flink application as a YARN job on a long-running cluster or on a transient cluster. On a long-running cluster, you can submit multiple Flink jobs to one Flink cluster running on Amazon EMR. If you run a Flink job on a transient cluster, your Amazon EMR cluster exists only for the time it takes to run the Flink application, so you are only charged for the resources and time used. You can submit a Flink job with the Amazon EMR `AddSteps` API operation, as a step argument to the `RunJobFlow` operation, and through the AWS CLI `add-steps` or `create-cluster` commands.

Start a Flink YARN application as a step on a long-running cluster

To start a Flink application that multiple clients can submit work to through YARN API operations, you need to either create a cluster or add a Flink application an existing cluster. For instructions on how to create a new cluster, see Creating a cluster with Flink. To start a YARN session on an existing cluster, use the following steps from the console, AWS CLI, or Java SDK.

> **Note**
>
> The `flink-yarn-session` command was added in Amazon EMR version 5.5.0 as a wrapper for the `yarn-session.sh` script to simplify execution. If you use an earlier version of Amazon EMR, substitute `bash -c "/usr/lib/flink/bin/yarn-session.sh -d"` for **Arguments** in the console or `Args`. in the AWS CLI command.

To submit a Flink job on an existing cluster from the console

Submit the Flink session with the `flink-yarn-session` command in an existing cluster.

1. Open the Amazon EMR console at https://console.aws.amazon.com/emr.
2. In the cluster list, select the cluster you previously launched.
3. In the cluster details page, choose **Steps**, **Add Step**.

4. Use the guidelines that follow to enter the parameters, and then choose **Add**.

Parameter	Description
Step type	**Custom JAR**
Name	A name to help you identify the step. For example, *<example-flink-step-name>*.
Jar location	**command-runner.jar**
Arguments	The `flink-yarn-session` command with arguments appropriate for your application. For example, *`flink-yarn-session -d`* starts a Flink session within your YARN cluster in a detached state (-d). See [YARN setup](#) in the latest Flink documentation for argument details.

To submit a Flink job on an existing cluster with the AWS CLI

- Use the `add-steps` command to add a Flink job to a long-running cluster. The following example command specifies `Args="flink-yarn-session", "-d"` to start a Flink session within your YARN cluster in a detached state (-d). See YARN setup in the latest Flink documentation for argument details.

```
aws emr add-steps --cluster-id <j-XXXXXXXX> --steps Type=CUSTOM_JAR,Name=<example-
flink-step-name>,Jar=command-runner.jar,Args="flink-yarn-session","-d"
```

Submit work to an existing Flink application on a long-running cluster

If you already have an existing Flink application on a long-running cluster, you can specify the cluster's Flink application ID in order to submit work to it. To obtain the application ID, run `yarn application -list` on the AWS CLI or through the YarnClient API operation:

```
$ yarn application -list
```

```
16/09/07 19:32:13 INFO client.RMProxy: Connecting to ResourceManager at
 ip-10-181-83-19.ec2.internal/10.181.83.19:8032
Total number of applications (application-types: [] and states: [SUBMITTED, ACCEPTED,
 RUNNING]):1
Application-Id     Application-Name     Application-Type     User     Queue     State
 Final-State     Progress     Tracking-URL
application_1473169569237_0002     Flink session with 14 TaskManagers (detached)
   Apache Flink      hadoop     default            RUNNING          UNDEFINED
 100% http://ip-10-136-154-194.ec2.internal:33089
```

The application ID for this Flink session is `application_1473169569237_0002`, which you can use to submit work to the application from the AWS CLI or an SDK.

Example SDK for Java

```
List<StepConfig> stepConfigs = new ArrayList<StepConfig>();

HadoopJarStepConfig flinkWordCountConf = new HadoopJarStepConfig()
    .withJar("command-runner.jar")
    .withArgs("flink", "run", "-m", "yarn-cluster", "-yid",
 "application_1473169569237_0002", "-yn", "2", "/usr/lib/flink/examples/streaming/
WordCount.jar",
      "--input", "s3://myBucket/pg11.txt", "--output", "s3://myBucket/alice2/");

StepConfig flinkRunWordCount = new StepConfig()
  .withName("Flink add a wordcount step")
  .withActionOnFailure("CONTINUE")
  .withHadoopJarStep(flinkWordCountConf);

stepConfigs.add(flinkRunWordCount);

AddJobFlowStepsResult res = emr.addJobFlowSteps(new AddJobFlowStepsRequest()
   .withJobFlowId("myClusterId")
   .withSteps(stepConfigs));
```

Example AWS CLI

```
aws emr add-steps --cluster-id <j-XXXXXXXX> \
--steps Type=CUSTOM_JAR,Name=Flink_Submit_To_Long_Running,Jar=command-runner.jar,\
Args="flink","run","-m","yarn-cluster","-yid","application_1473169569237_0002",\
"/usr/lib/flink/examples/streaming/WordCount.jar",\
"--input","s3://myBucket/pg11.txt","--output","s3://myBucket/alice2/" \
```

```
--region <region-code>
```

Submit a transient Flink job

The following examples launch a transient cluster that runs a Flink job and then terminates on completion.

Example SDK for Java

```
import java.util.ArrayList;
import java.util.List;
import com.amazonaws.AmazonClientException;
import com.amazonaws.auth.AWSCredentials;
import com.amazonaws.auth.AWSStaticCredentialsProvider;
import com.amazonaws.auth.profile.ProfileCredentialsProvider;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduce;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduceClientBuilder;
import com.amazonaws.services.elasticmapreduce.model.*;

public class Main_test {

 public static void main(String[] args) {
  AWSCredentials credentials_profile = null;
  try {
   credentials_profile = new ProfileCredentialsProvider("default").getCredentials();
  } catch (Exception e) {
   throw new AmazonClientException(
     "Cannot load credentials from .aws/credentials file. " +
       "Make sure that the credentials file exists and the profile name is specified
 within it.",
     e);
  }

  AmazonElasticMapReduce emr = AmazonElasticMapReduceClientBuilder.standard()
    .withCredentials(new AWSStaticCredentialsProvider(credentials_profile))
    .withRegion(Regions.US_WEST_1)
    .build();

  List<StepConfig> stepConfigs = new ArrayList<StepConfig>();
  HadoopJarStepConfig flinkWordCountConf = new HadoopJarStepConfig()
    .withJar("command-runner.jar")
    .withArgs("bash", "-c", "flink", "run", "-m", "yarn-cluster", "-yn", "2",
```

```
      "/usr/lib/flink/examples/streaming/WordCount.jar", "--input", "s3://path/to/
input-file.txt", "--output",
      "s3://path/to/output/");

  StepConfig flinkRunWordCountStep = new StepConfig()
    .withName("Flink add a wordcount step and terminate")
    .withActionOnFailure("CONTINUE")
    .withHadoopJarStep(flinkWordCountConf);

  stepConfigs.add(flinkRunWordCountStep);

  Application flink = new Application().withName("Flink");

  RunJobFlowRequest request = new RunJobFlowRequest()
    .withName("flink-transient")
    .withReleaseLabel("emr-5.20.0")
    .withApplications(flink)
    .withServiceRole("EMR_DefaultRole")
    .withJobFlowRole("EMR_EC2_DefaultRole")
    .withLogUri("s3://path/to/my/logfiles")
    .withInstances(new JobFlowInstancesConfig()
      .withEc2KeyName("myEc2Key")
      .withEc2SubnetId("subnet-12ab3c45")
      .withInstanceCount(3)
      .withKeepJobFlowAliveWhenNoSteps(false)
      .withMasterInstanceType("m4.large")
      .withSlaveInstanceType("m4.large"))
    .withSteps(stepConfigs);

  RunJobFlowResult result = emr.runJobFlow(request);
  System.out.println("The cluster ID is " + result.toString());

 }

}
```

Example AWS CLI

Use the `create-cluster` subcommand to create a transient cluster that terminates when the Flink job completes:

```
aws emr create-cluster --release-label emr-5.2.1 \
--name "Flink_Transient" \
```

```
--applications Name=Flink \
--configurations file://./configurations.json \
--region us-east-1 \
--log-uri s3://myLogUri \
--auto-terminate
--instance-type m5.xlarge \
--instance-count 2 \
--service-role EMR_DefaultRole_V2 \
--ec2-attributes KeyName=<YourKeyName>,InstanceProfile=EMR_EC2_DefaultRole \
--steps Type=CUSTOM_JAR,Jar=command-runner.jar,Name=Flink_Long_Running_Session,\
Args="bash","-c","\"flink run -m yarn-cluster /usr/lib/flink/examples/streaming/
WordCount.jar
--input s3://myBucket/pg11.txt --output s3://myBucket/alice/""
```

Use the Scala shell

The Flink Scala shell for EMR clusters is only configured to start new YARN sessions. You can use the Scala shell by following the procedure below.

Use the Flink Scala shell on the primary node

1. Log in to the primary node with SSH as described in Connect to the primary node with SSH.
2. Type the following to start a shell:

   In Amazon EMR version 5.5.0 and later, you can use the following command to start a Yarn cluster for the Scala Shell with one TaskManager.

   ```
   % flink-scala-shell yarn 1
   ```

   In earlier versions of Amazon EMR, use:

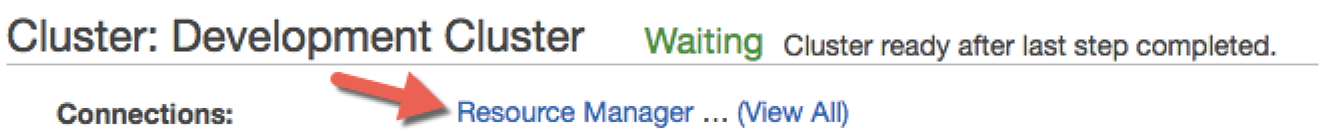
   ```
   % /usr/lib/flink/bin/start-scala-shell.sh yarn 1
   ```

   This starts the Flink Scala shell so you can interactively use Flink. Just as with other interfaces and options, you can scale the `-n` option value used in the example based on the number of tasks you want to run from the shell.

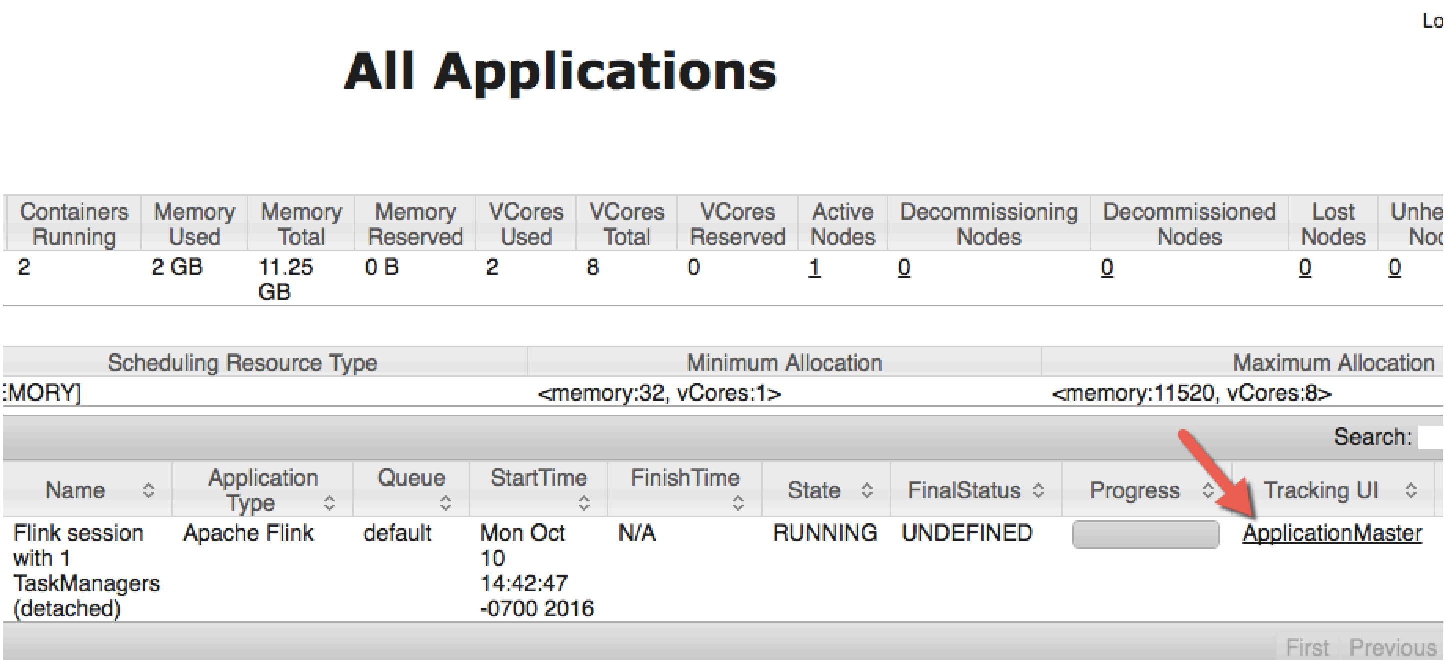
   For more information, see Scala REPL in the official Apache Flink documentation.

Finding the Flink web interface

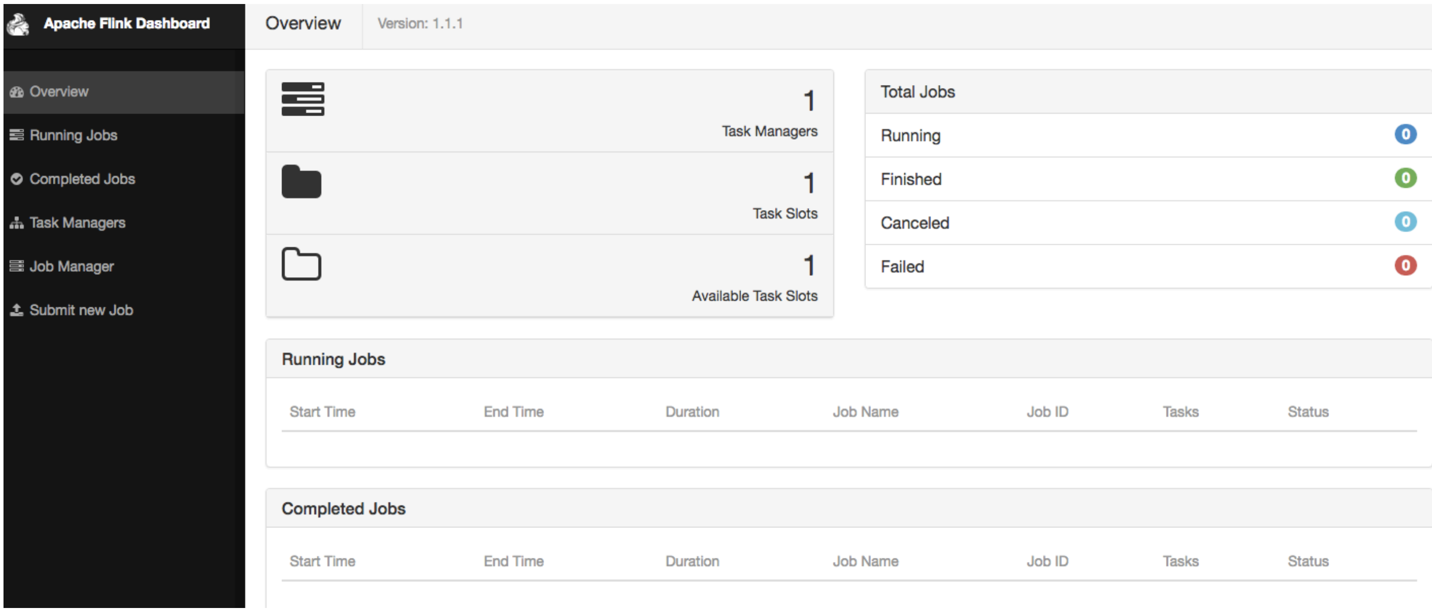
The Application Master that belongs to the Flink application hosts the Flink web interface. It is an alternative way to submit a JAR as a job or to view the current status of other jobs. The Flink web interface is active as long as you have a Flink session running. If you have a long-running YARN job already active, you can follow the instructions in the [Connect to the primary node with SSH](#) topic in the *Amazon EMR Management Guide* to connect to the YARN ResourceManager. For example, if you've set up an SSH tunnel and have activated a proxy in your browser, you choose the ResourceManager connection under **Connections** in your EMR cluster details page.



After you find the ResourceManager, select the YARN application that's hosting a Flink session. Choose the link under the **Tracking UI** column.



In the Flink web interface, you can view configuration, submit your own custom JAR as a job, or monitor jobs in progress.



Flink autoscaler

Overview

Amazon EMR releases 6.15.0 and higher support *Flink autoscaler*. The job autoscaler functionality collects metrics from running Flink streaming jobs, and automatically scales the individual job vertexes. This reduces the backpressure and satisfies the utilization target that you set.

For more information, see the Autoscaler section of the *Apache Flink Kubernetes Operator* documentation.

Considerations

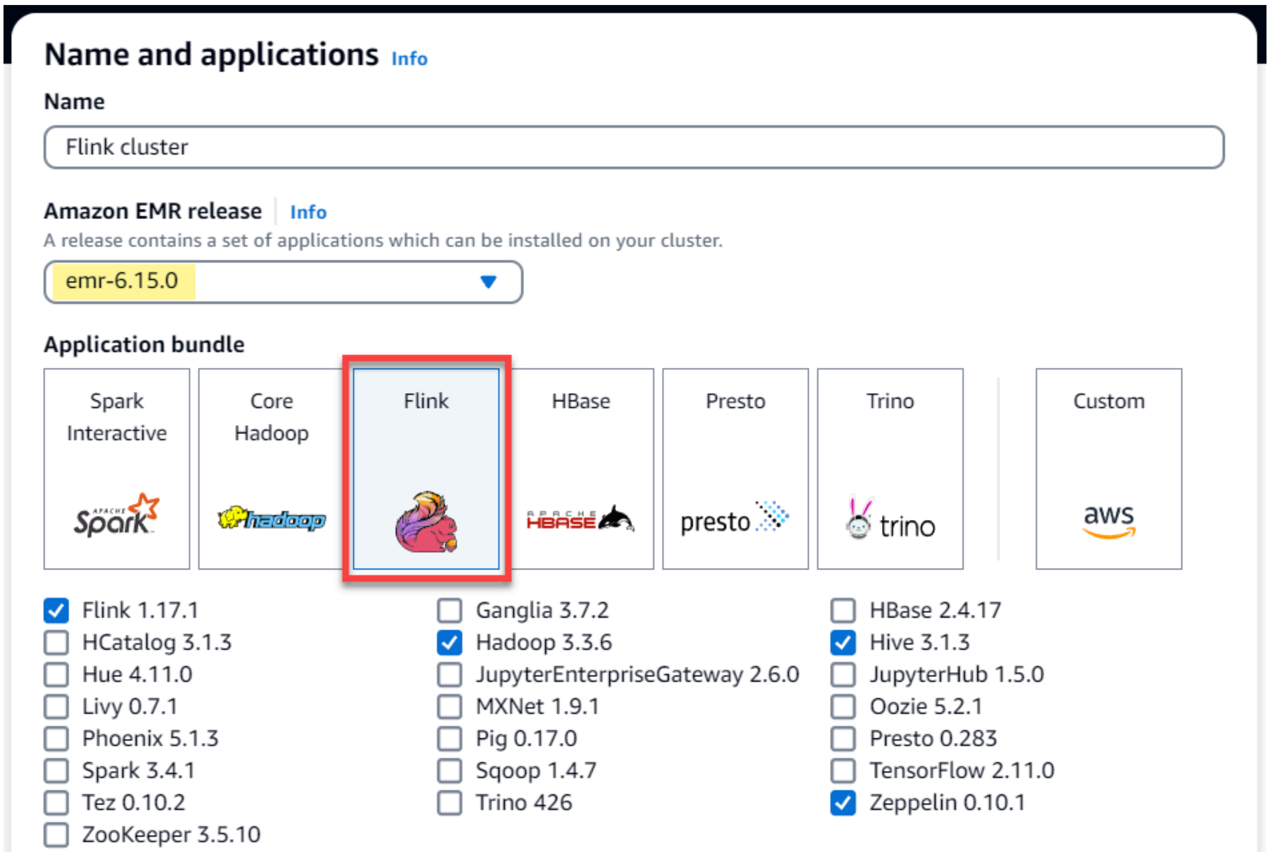
- Flink autoscaler is supported with Amazon EMR 6.15.0 and higher.
- Flink autoscaler is supported only for streaming jobs.
- Only adaptive scheduler is supported. The default scheduler is not supported.
- We recommend that you enable cluster scaling to allow dynamic resource provision. Amazon EMR managed scaling is preferred `because the metric evaluation occurs every 5–10 seconds. At this interval, your cluster can more readily adjust to the change in the required cluster resources.

Enable autoscaler

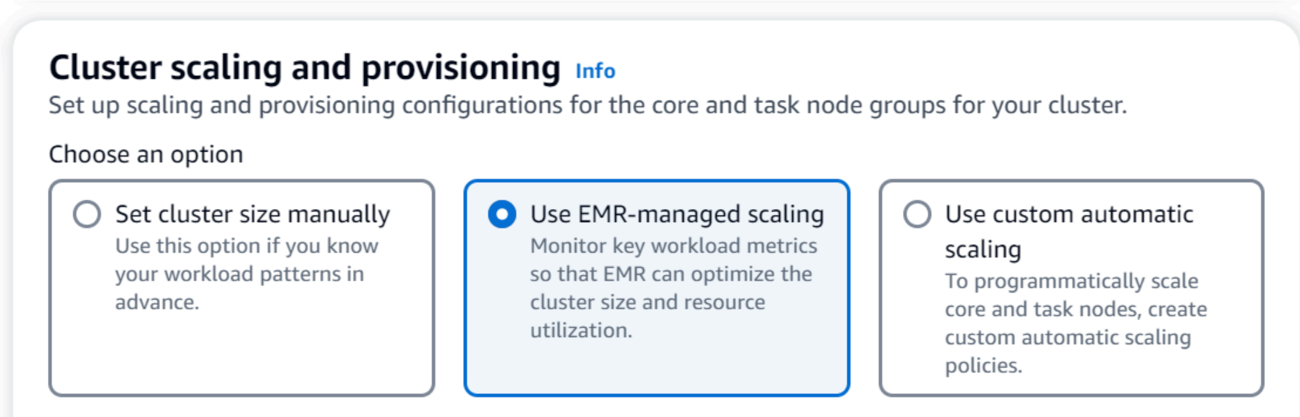
Use the following steps to enable the Flink autoscaler when you create an Amazon EMR on EC2 cluster.

1. In the Amazon EMR console, create a new EMR cluster:

   a. Choose Amazon EMR release `emr-6.15.0` or higher. Select the **Flink** application bundle, and select any other applications that you might want to include on your cluster.



   b. For the **Cluster scaling and provisioning** option, select **Use EMR-managed scaling**.



2. In the **Software settings** section, enter the following configuration to enable Flink autoscaler. For testing scenarios, set the decision interval, metrics window interval, and stabilization interval to a lower value so that the job immediately makes a scaling decision for easier verification.

```
[
  {
    "Classification": "flink-conf",
    "Properties": {
      "job.autoscaler.enabled": "true",
      "jobmanager.scheduler": "adaptive",
      "job.autoscaler.stabilization.interval": "60s",
      "job.autoscaler.metrics.window": "60s",
      "job.autoscaler.decision.interval": "10s",
      "job.autoscaler.debug.logs.interval": "60s"
    }
  }
]
```

3. Select or configure any other settings as you prefer them, and create the Flink autoscaler-enabled cluster.

Autoscaler configurations

This section covers most of the configurations that you can change based on your specific needs.

> **Note**
>
> With time-based configurations like `time`, `interval` and `window` settings, the default unit when no unit is specified is milliseconds. So a value of `30` with no suffix equals

30 milliseconds. For other units of time, include the appropriate suffix of s for *seconds*, m for *minutes*, or h for *hours*.

Topics

- Autoscaler loop configurations
- Metrics aggregation and history configurations
- Job vertex level configurations
- Backlog processing configurations
- Scale operation configurations

Autoscaler loop configurations

Autoscaler fetches the job vertex level metrics for every few configurable time interval, converts them into scale actionables, estimates new job vertex parallelism, and recommends it to job scheduler. Metrics are collected only after the job restart time and cluster stabilization interval.

Config key	Default value	Description	Example values
`job.autoscaler.enabled`	`false`	Enable autoscaling on your Flink cluster.	`true`, `false`
`job.autoscaler.decision.interval`	`60s`	Autoscaler decision interval.	`30` (default unit is milliseconds), `5m`, `1h`
`job.autoscaler.restart.time`	`3m`	Expected restart time to be used until the operator can determine it reliably from history.	`30` (default unit is milliseconds), `5m`, `1h`
`job.autoscaler.stabilization.interval`	`300s`	Stabilization period in which no new scaling will be executed.	`30` (default unit is

Config key	Default value	Description	Example values
			milliseconds), 5m, 1h
job.autoscaler.debug.logs.interval	300s	Autoscaler debug logs interval.	30 (default unit is milliseconds), 5m, 1h

Metrics aggregation and history configurations

Autoscaler fetches the metrics, aggregates them over time based sliding window and these are evaluated into scaling decisions. The scaling decision history for each job vertex are utilised to estimate new parallelism. These have both age based expiry as well as history size (at-least 1).

Config key	Default value	Description	Example values
job.autoscaler.metrics.window	600s	Scaling metrics aggregation window size.	30 (default unit is milliseconds), 5m, 1h
job.autoscaler.history.max.count	3	Maximum number of past scaling decisions to retain per vertex.	1 to Integer.MAX_VALUE
job.autoscaler.history.max.age	24h	Minimum number of past scaling decisions to retain per vertex.	30 (default unit is milliseconds), 5m, 1h

Job vertex level configurations

The parallelism of each job vertex is modified on the basis of target utilisation and bounded by the min-max parallelism limits. It’s not recommended to set target utilisation close to 100% (i.e value of 1) and the utilisation boundary works as a buffer to handle the intermediate load fluctuations.

Config key	Default value	Description	Example values
`job.autoscaler.target.utilization`	`0.7`	Target vertex utilization.	`0` - `1`
`job.autoscaler.target.utilization.boundary`	`0.4`	Target vertex utilization boundary. Scaling won't be performed if the current processing rate is within [`target_rate / (target_utilization - boundary)` , and (`target_rate / (target_utilization + boundary)`]	`0` - `1`
`job.autoscaler.vertex.min-parallelism`	`1`	The minimum parallelism that the autoscaler can use.	`0` - `200`
`job.autoscaler.vertex.max-parallelism`	`200`	The maximum parallelism the autoscaler can use. Note that this limit will be ignored if it is higher than the max parallelism configured in the Flink config or directly on each operator.	`0` - `200`

Backlog processing configurations

The job vertex needs extra resources to handle the pending events, or backlogs, that accumulate during the scale operation time period. This is also referred as the `catch-up` duration. If the time to process backlog exceeds the configured `lag -threshold` value, the job vertex target utilization increases to max level. This helps prevent unnecessary scaling operations while the backlog processes.

Config key	Default value	Description	Example values
`job.autoscaler.backlog-processing.lag-threshold`	`5m`	Lag threshold which will prevent unnecessary scalings while removing the pending messages responsible for the lag.	`30` (default unit is milliseconds), `5m`, `1h`
`job.autoscaler.catch-up.duration`	`15m`	The target duration for fully processing any backlog after a scaling operation. Set to 0 to disable backlog based scaling.	`30` (default unit is milliseconds), `5m`, `1h`

Scale operation configurations

Autoscaler doesn't perform scale down operation immediately after a scale up operation within grace time period. This prevents un-necessary cycle of scale up-down-up-down operations caused by temporary load fluctuations.

We can use the scale down operation ratio to gradually decrease the parallelism and release resources to cater for temporary load spike. It also helps to prevent un-necessary minor scale up operation post major scale down operation.

We can detect an in-effective scale up operation based past job vertex scaling decision history to prevent further parallelism change.

Config key	Default value	Description	Example values
job.autoscaler.scale-up.grace-period	1h	Duration in which no scale down of a vertex is allowed after it has been scaled up.	30 (default unit is milliseconds), 5m, 1h
job.autoscaler.scale-down.max-factor	0.6	Max scale down factor. A value of 1 means no limit on scale down; 0.6 means job can only be scaled down with 60% of the original parallelism.	0 - 1
job.autoscaler.scale-up.max-factor	100000.	Maximum scale up ratio. A value of 2.0 means job can only be scaled up with 200% of the current parallelism.	0 - Integer.MAX_VALUE
job.autoscaler.scaling.effectiveness.detection.enabled	false	Whether to enable detection of ineffective scaling operations and allowing the autoscaler to block further scale ups.	true, false

Optimizing job restart times for task recovery and scaling operations

When a task fails or when a scaling operation occurs, Flink attempts to re-execute the task from the last completed checkpoint. The restart process could take a minute or longer to execute, depending on the size of the checkpoint state and the number of parallel tasks. During the restart period, backlog tasks can accumulate for the job. There are some ways though, that Flink optimizes the speed of recovery and restart of execution graphs to improve job stability.

This page describes some of the ways that Amazon EMR Flink can improve the job restart time during task recovery or scaling operations.

Topics

Task-local recovery

> **Note**
>
> Task-local recovery is supported with Amazon EMR 6.0.0 and higher.

With Flink checkpoints, each task produces a snapshot of its state that Flink writes to distributed storage like Amazon S3. In cases of recovery, the tasks restore their state from the distributed storage. Distributed storage provides fault tolerance and can redistribute the state during rescaling because it's accessible to all nodes.

However, a remote distributed store also has a disadvantage: all tasks must read their state from a remote location over the network. This can result in long recovery times for large states during task recovery or scaling operations.

This problem of long recovery time is solved by *task-local recovery*. Tasks write their state on checkpoint into a secondary storage that is local to the task, such as on a local disk. They also store their state in the primary storage, or Amazon S3 in our case. During recovery, the scheduler schedules the tasks on the same task manager where the tasks ran earlier so that they can recover from the local state store instead of reading from the remote state store. For more information, see Task-Local Recovery in the *Apache Flink Documentation*.

Our benchmark tests with sample jobs have shown that the recovery time has been reduced from minutes to a few seconds with task-local recovery enabled.

To enable task-local recovery, set the following configurations in your `flink-conf.yaml` file. Specify the checkpointing interval value in milliseconds.

```
state.backend.local-recovery: true
state.backend: hasmap or rocksdb
state.checkpoints.dir: s3://storage-location-bucket-path/checkpoint
execution.checkpointing.interval: 15000
```

Generic log-based incremental checkpoint

Note

Generic log-based incremental checkpointing is supported with Amazon EMR 6.10.0 and higher.

Generic log-based incremental checkpointing was added in Flink 1.16 to improve the speed of checkpoints. A faster checkpoint interval often results in a reduction of recovery work because fewer events need to be reprocessed after recovery. For more information, see Improving speed and stability of checkpointing with generic log-based incremental checkpoints on the *Apache Flink Blog*.

With sample jobs, our benchmark tests have shown that the checkpoint time reduced from minutes to a few seconds with the generic log-based incremental checkpoint.

To enable generic log-based incremental checkpoints, set the following configurations in your `flink-conf.yaml` file. Specify the checkpointing interval value in milliseconds.

```
state.backend.changelog.enabled: true
state.backend.changelog.storage: filesystem
dstl.dfs.base-path: s3://bucket-path/changelog
state.backend.local-recovery: true
state.backend: rocksdb
state.checkpoints.dir: s3://bucket-path/checkpoint
execution.checkpointing.interval: 15000
```

Fine-grained recovery

> **Note**
>
> Fine-grained recovery support for the default scheduler is available with Amazon EMR 6.0.0 and higher. Fine-grained recovery support in the adaptive scheduler is available with Amazon EMR 6.15.0 and higher.

When a task fails during execution, Flink resets the entire execution graph and triggers complete re-execution from the last completed checkpoint. This is more expensive than just re-executing the failed tasks. Fine-grained recovery restarts only the pipeline-connected component of the failed task. In the following example, the job graph has 5 vertices (`A` to `E`). All connections between the vertices are pipelined with pointwise distribution, and the `parallelism.default` for the job is set to `2`.

```
A # B # C # D # E
```

For this example, there are a total of 10 tasks running. The first pipeline (`a1` to `e1`) runs on a TaskManager (`TM1`), and the second pipeline (`a2` to `e2`) runs on another TaskManager (`TM2`).

```
a1 # b1 # c1 # d1 # e1
a2 # b2 # c2 # d2 # e2
```

There are two pipelined connected components: `a1 # e1`, and `a2 # e2`. If either `TM1` or `TM2` fails, the failure impacts only the 5 tasks in the pipeline where the TaskManager was running. The restart strategy only starts the affected pipelined component.

Fine-grained recovery works only with perfectly parallel Flink jobs. It's not supported with `keyBy()` or `redistribute()` operations. For more information, see [FLIP-1: Fine Grained Recovery from Task Failures](#) in the *Flink Improvement Proposal* Jira project.

To enable fine-grained recovery, set the following configurations in your `flink-conf.yaml` file.

```
jobmanager.execution.failover-strategy: region
restart-strategy: exponential-delay or fixed-delay
```

Combined restart mechanism in adaptive scheduler

> **Note**
>
> The combined restart mechanism in adaptive scheduler is supported with Amazon EMR 6.15.0 and higher.

Adaptive scheduler can adjust the parallelism of the job based on available slots. It automatically reduces the parallelism if not enough slots are available to fit the configured job parallelism. If new slots become available, the job is scaled up again to the configured job parallelism. An adaptive scheduler avoids downtime on the job when there are not enough resources available. This is the supported scheduler for Flink Autoscaler. We recommend adaptive scheduler with Amazon EMR Flink for these reasons. However, adaptive schedulers might do multiple restarts within a short period of time, one restart for every new resource added. This could lead to a performance drop in the job.

With Amazon EMR 6.15.0 and higher, Flink has a combined restart mechanism in adaptive scheduler that opens a restart window when the first resource is added, and then waits until the configured window interval of the default 1 minute. It performs a single restart when there are sufficient resources available to run the job with configured parallelism or when the interval times out.

With sample jobs, our benchmark tests have shown that this feature processes 10% of records more than the default behavior when you use adaptive scheduler and Flink autoscaler.

To enable the combined restart mechanism, set the following configurations in your `flink-conf.yaml` file.

```
jobmanager.adaptive-scheduler.combined-restart.enabled: true
jobmanager.adaptive-scheduler.combined-restart.window-interval: 1m
```

Working with Flink jobs from Zeppelin in Amazon EMR

Introduction

Amazon EMR releases 6.10.0 and higher support [Apache Zeppelin](https://zeppelin.apache.org/) integration with Apache Flink. You can interactively submit Flink jobs through Zeppelin notebooks. With the Flink interpreter, you can execute Flink queries, define Flink streaming and batch jobs, and visualize the output within Zeppelin notebooks. The Flink interpreter is built on top of the Flink REST API. This lets you access and manipulate Flink jobs from within the Zeppelin environment to perform real-time data processing and analysis.

There are four sub-interpreters in Flink interpreter. They serve different purposes, but are all in the the JVM and share the same pre-configured entry points to Flink (`ExecutionEnviroment`, `StreamExecutionEnvironment`, `BatchTableEnvironment`, `StreamTableEnvironment`). The interpreters are as follows:

- `%flink` – Creates `ExecutionEnvironment`, `StreamExecutionEnvironment`, `BatchTableEnvironment`, `StreamTableEnvironment`, and provides a Scala environment
- `%flink.pyflink` – Provides a Python environment
- `%flink.ssql` – Provides a streaming SQL environment
- `%flink.bsql` – Provides a batch SQL environment

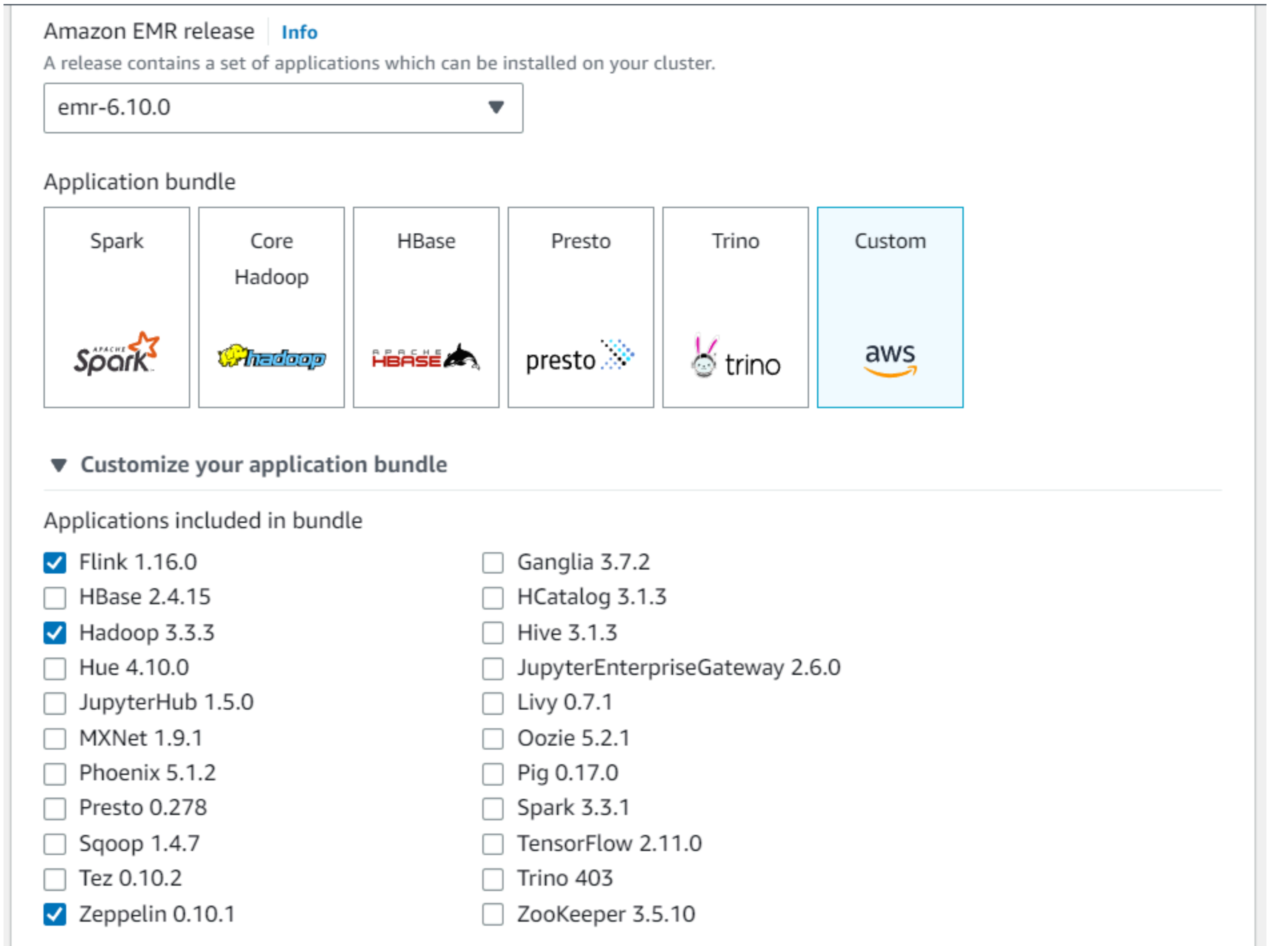
Prerequisites

- Zeppelin integration with Flink is supported for clusters created with Amazon EMR 6.10.0 and higher.
- To view web interfaces that are hosted on EMR clusters as required for these steps, you must configure an SSH tunnel to allow inbound access. For more information, see [Configure proxy settings to view websites hosted on the primary node](#).

Configure Zeppelin-Flink on an EMR cluster

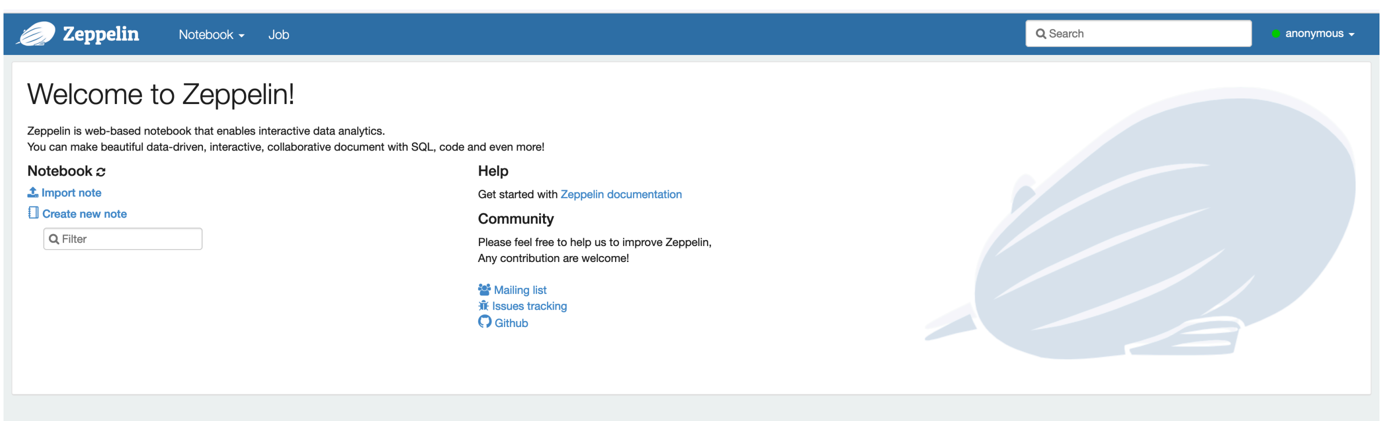
Use the following steps to configure Apache Flink on Apache Zeppelin to run on an EMR cluster:

1. Create a new cluster from the Amazon EMR console. Select emr-6.10.0 or higher for the Amazon EMR release. Then, choose to customize your application bundle with the Custom option. Include at least Flink, Hadoop, and Zeppelin in your bundle.

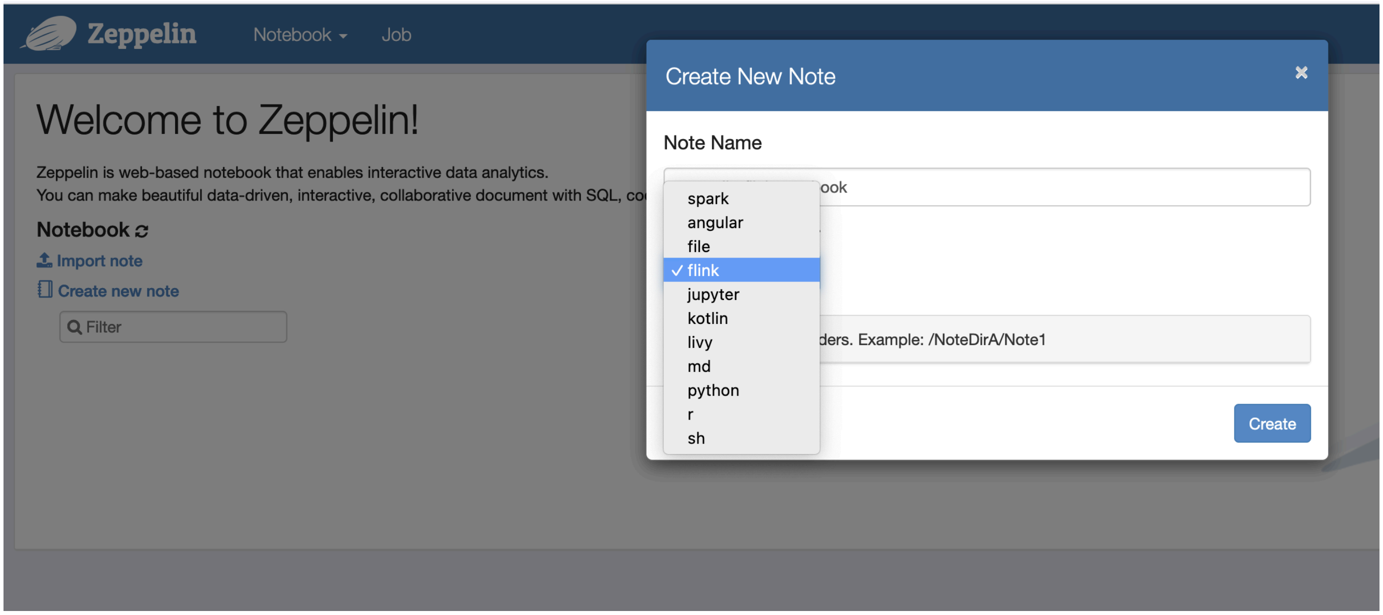


2. Create the rest of your cluster with the settings that you prefer.

3. Once your cluster is running, select the cluster in the console to view its details and open the Applications tab. Select Zeppelin from the Application user interfaces section to open the Zeppelin web interface. Be sure that you've set up access to the Zeppelin web interface with an SSH tunnel to the primary node and a proxy connection as described in the Prerequisites.



4. Now, you can create a new note in a Zeppelin notebook with Flink as the default interpreter.



5. Refer to the following code examples that demonstrate how to run Flink jobs from a Zeppelin notebook.

Run Flink jobs with Zeppelin-Flink on an EMR cluster

- Example 1, Flink Scala

  a) Batch WordCount Example (SCALA)

```
%flink

val data = benv.fromElements("hello world", "hello flink", "hello hadoop")
data.flatMap(line => line.split("\\s"))
             .map(w => (w, 1))
             .groupBy(0)
             .sum(1)
             .print()
```

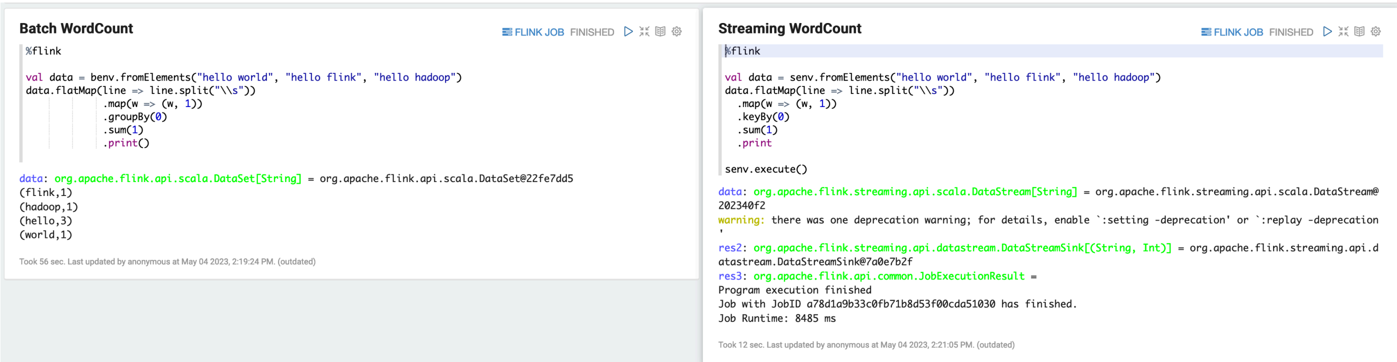
  b) Streaming WordCount Example (SCALA)

```
%flink

val data = senv.fromElements("hello world", "hello flink", "hello hadoop")
data.flatMap(line => line.split("\\s"))
  .map(w => (w, 1))
```

```
  .keyBy(0)
  .sum(1)
  .print

senv.execute()
```



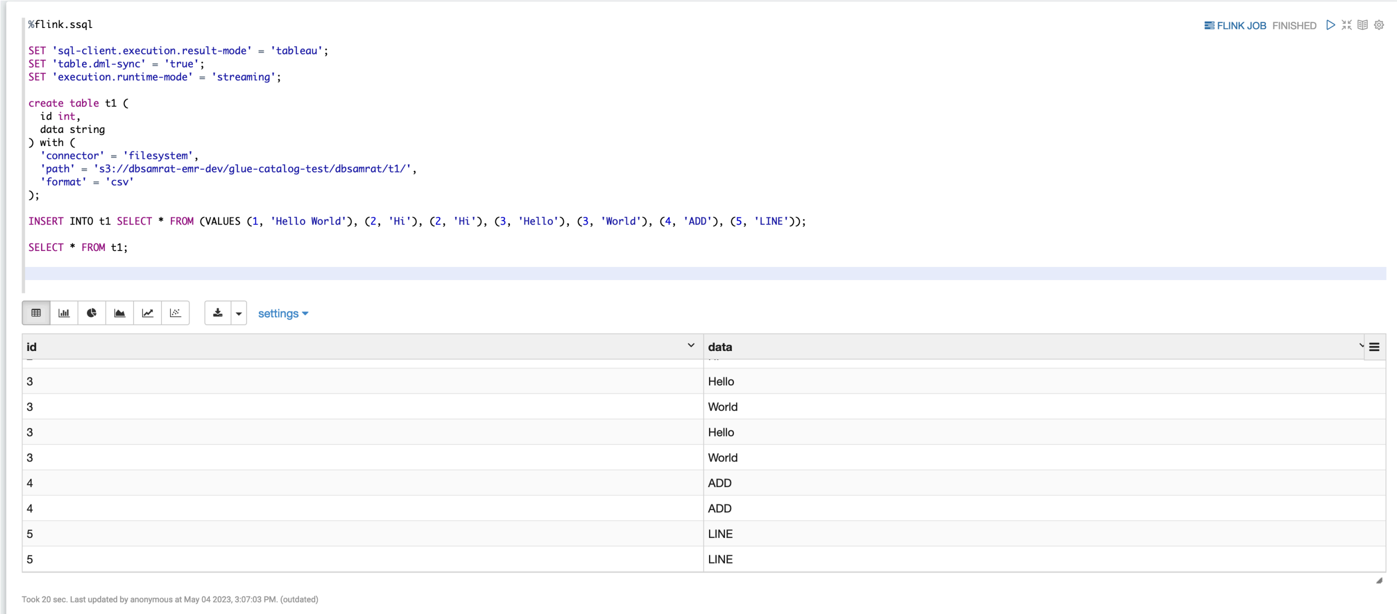
- Example 2, Flink Streaming SQL

```
%flink.ssql
SET 'sql-client.execution.result-mode' = 'tableau';
SET 'table.dml-sync' = 'true';
SET 'execution.runtime-mode' = 'streaming';

create table dummy_table (
  id int,
  data string
) with (
  'connector' = 'filesystem',
  'path' = 's3://s3-bucket/dummy_table',
  'format' = 'csv'
);

INSERT INTO dummy_table SELECT * FROM (VALUES (1, 'Hello World'), (2, 'Hi'), (2,
 'Hi'), (3, 'Hello'), (3, 'World'), (4, 'ADD'), (5, 'LINE'));

SELECT * FROM dummy_table;
```



- Example 3, Pyflink. Note that you must upload your own sample text file named `word.txt` to your S3 bucket.

```
%flink.pyflink

import argparse
import logging
import sys

from pyflink.common import Row
from pyflink.table import (EnvironmentSettings, TableEnvironment, TableDescriptor,
 Schema,
                           DataTypes, FormatDescriptor)
from pyflink.table.expressions import lit, col
from pyflink.table.udf import udtf

def word_count(input_path, output_path):
    t_env = TableEnvironment.create(EnvironmentSettings.in_streaming_mode())
    # write all the data to one file
    t_env.get_config().set("parallelism.default", "1")

    # define the source
    if input_path is not None:
        t_env.create_temporary_table(
            'source',
            TableDescriptor.for_connector('filesystem')
```

```
                            .schema(Schema.new_builder()
                                    .column('word', DataTypes.STRING())
                                    .build())
                            .option('path', input_path)
                            .format('csv')
                            .build())
        tab = t_env.from_path('source')
    else:
        print("Executing word_count example with default input data set.")
        print("Use --input to specify file input.")
        tab = t_env.from_elements(map(lambda i: (i,), word_count_data),
                                   DataTypes.ROW([DataTypes.FIELD('line', 
DataTypes.STRING())]))

    # define the sink
    if output_path is not None:
        t_env.create_temporary_table(
            'sink',
            TableDescriptor.for_connector('filesystem')
                            .schema(Schema.new_builder()
                                    .column('word', DataTypes.STRING())
                                    .column('count', DataTypes.BIGINT())
                                    .build())
                            .option('path', output_path)
                            .format(FormatDescriptor.for_format('canal-json')
                                    .build())
                            .build())
    else:
        print("Printing result to stdout. Use --output to specify output path.")
        t_env.create_temporary_table(
            'sink',
            TableDescriptor.for_connector('print')
                            .schema(Schema.new_builder()
                                    .column('word', DataTypes.STRING())
                                    .column('count', DataTypes.BIGINT())
                                    .build())
                            .build())

    @udtf(result_types=[DataTypes.STRING()])
    def split(line: Row):
        for s in line[0].split():
            yield Row(s)

    # compute word count
```

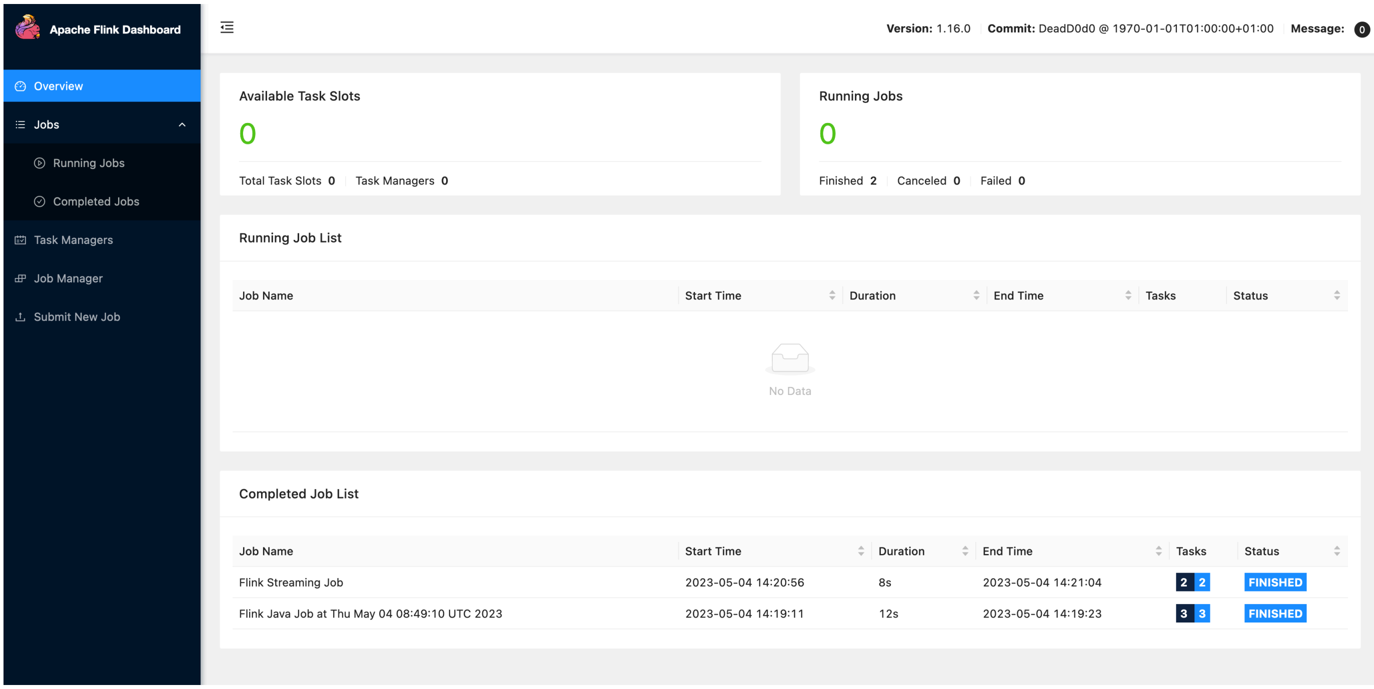
```
    tab.flat_map(split).alias('word') \
       .group_by(col('word')) \
       .select(col('word'), lit(1).count) \
       .execute_insert('sink') \
       .wait()


logging.basicConfig(stream=sys.stdout, level=logging.INFO, format="%(message)s")


word_count("s3://s3_bucket/word.txt", "s3://s3_bucket/demo_output.txt")
```

1. Choose **FLINK JOB** in the Zeppelin UI to access and view the Flink Web UI.

2. Choosing **FLINK JOB** routes to the Flink Web Console in another tab of your browser.



Flink release history

The following table lists the version of Flink included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in [Amazon EMR 7.x release versions](), [Amazon EMR 6.x release versions](), or [Amazon EMR 5.x release versions]().

Flink version information

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-7.2.0	1.18.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server,

Amazon EMR Release label	Flink Version	Components installed with Flink
		flink-client, flink-jobmanager-config, hudi, delta
emr-5.36.2	1.14.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config
emr-7.1.0	1.18.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-7.0.0	1.18.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors
emr-6.15.0	1.17.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-6.14.0	1.17.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors
emr-6.13.0	1.17.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-6.12.0	1.17.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors
emr-6.11.1	1.16.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-6.11.0	1.16.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors
emr-6.10.1	1.16.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-6.10.0	1.16.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-6.9.1	1.15.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-6.9.0	1.15.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-6.8.1	1.15.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-6.8.0	1.15.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-6.7.0	1.14.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.36.1	1.14.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config
emr-5.36.0	1.14.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-6.6.0	1.14.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-5.35.0	1.14.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-6.5.0	1.14.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-6.4.0	1.13.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-6.3.1	1.12.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config
emr-6.3.0	1.12.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-6.2.1	1.11.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config
emr-6.2.0	1.11.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config
emr-6.1.1	1.11.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-6.1.0	1.11.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.34.0	1.13.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config
emr-5.33.1	1.12.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.33.0	1.12.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config
emr-5.32.1	1.11.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.32.0	1.11.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config
emr-5.31.1	1.11.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.31.0	1.11.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config
emr-5.30.2	1.10.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.30.1	1.10.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.30.0	1.10.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.29.0	1.9.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.28.1	1.9.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.28.0	1.9.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.27.1	1.8.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.27.0	1.8.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.26.0	1.8.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.25.0	1.8.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.24.1	1.8.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.24.0	1.8.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.23.1	1.7.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.23.0	1.7.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.22.0	1.7.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.21.2	1.7.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.21.1	1.7.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.21.0	1.7.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.20.1	1.6.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.20.0	1.6.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.19.1	1.6.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.19.0	1.6.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.18.1	1.6.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.18.0	1.6.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.17.2	1.5.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.17.1	1.5.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.17.0	1.5.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.16.1	1.5.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.16.0	1.5.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.15.1	1.4.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.15.0	1.4.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.14.2	1.4.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.14.1	1.4.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.14.0	1.4.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.13.1	1.4.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.13.0	1.4.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.12.3	1.4.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.12.2	1.4.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.12.1	1.4.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.12.0	1.4.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.11.4	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.11.3	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.11.2	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.11.1	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.11.0	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.10.1	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.10.0	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.9.1	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.9.0	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.8.3	1.3.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.8.2	1.3.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.8.1	1.3.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.8.0	1.3.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.7.1	1.3.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.7.0	1.3.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.6.1	1.2.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.6.0	1.2.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.5.4	1.2.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client
emr-5.5.3	1.2.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.5.2	1.2.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client
emr-5.5.1	1.2.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client
emr-5.5.0	1.2.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.4.1	1.2.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client
emr-5.4.0	1.2.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client
emr-5.3.2	1.1.4	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.3.1	1.1.4	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client
emr-5.3.0	1.1.4	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client
emr-5.2.3	1.1.3	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.2.2	1.1.3	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client
emr-5.2.1	1.1.3	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client
emr-5.2.0	1.1.3	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client

Amazon EMR Release label	Flink Version	Components installed with Flink
emr-5.1.1	1.1.3	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client
emr-5.1.0	1.1.3	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client

Flink release notes by version

See the following sections for the full release notes.

Amazon EMR 7.2.0 - Flink release notes

Type	Description
Improvement	Support adding custom labels to Kubernetes service per Flink job via configuration `kubernetes.service.labels` .

Ganglia

> **Note**
>
> The last release of Amazon EMR to include Ganglia was Amazon EMR 6.15.0. To monitor your cluster, releases higher than 6.15.0 include the Amazon CloudWatch agent.

The Ganglia open source project is a scalable, distributed system designed to monitor clusters and grids while minimizing the impact on their performance. When you enable Ganglia on your cluster, you can generate reports and view the performance of the cluster as a whole, as well as inspect the performance of individual node instances. Ganglia is also configured to ingest and visualize Hadoop and Spark metrics. For more information about the Ganglia open-source project, go to http://ganglia.info/.

When you view the Ganglia web UI in a browser, you see an overview of the cluster's performance, with graphs detailing the load, memory usage, CPU utilization, and network traffic of the cluster. Below the cluster statistics are graphs for each individual server in the cluster.

The following table lists the version of Ganglia included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with Ganglia.

For the version of components installed with Ganglia in this release, see Release 6.15.0 Component Versions.

Ganglia version information for emr-6.15.0

Amazon EMR Release Label	Ganglia Version	Components Installed With Ganglia
emr-6.15.0	Ganglia 3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-

Amazon EMR Release Label	Ganglia Version	Components Installed With Ganglia
		yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

The following table lists the version of Ganglia included in the latest release of the Amazon EMR 5.x series, along with the components that Amazon EMR installs with Ganglia.

For the version of components installed with Ganglia in this release, see Release 5.36.2 Component Versions.

Ganglia version information for emr-5.36.2

Amazon EMR Release Label	Ganglia Version	Components Installed With Ganglia
emr-5.36.2	Ganglia 3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Topics

- Create a cluster with Ganglia
- View Ganglia metrics
- Hadoop and Spark metrics in Ganglia
- Ganglia release history

Create a cluster with Ganglia

Note

The last release of Amazon EMR to include Ganglia was Amazon EMR 6.15.0. To monitor your cluster, releases higher than 6.15.0 include the Amazon CloudWatch agent.

To create a cluster with Ganglia using the console

1. Navigate to the new Amazon EMR console and select **Switch to the old console** from the side navigation. For more information on what to expect when you switch to the old console, see Using the old console.
2. Choose **Create cluster**.
3. In **Software configuration**, choose either **All Applications**, **Core Hadoop**, or **Spark**.
4. Proceed with creating the cluster with configurations as appropriate.

To add Ganglia to a cluster using the AWS CLI

In the AWS CLI, you can add Ganglia to a cluster by using `create-cluster` with the `--applications` parameter. If you specify only Ganglia using the `--applications` parameter, Ganglia is the only application installed.

- Type the following command to add Ganglia when you create a cluster and replace *myKey* with the name of your EC2 key pair.

  Note

  Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --name "Spark cluster with Ganglia" --release-label
 emr-6.15.0; \
--applications Name=Spark Name=Ganglia \
--ec2-attributes KeyName=myKey --instance-type m5.xlarge \
--instance-count 3 --use-default-roles
```

When you specify the instance count without using the `--instance-groups` parameter, a single master node is launched, and the remaining instances are launched as core nodes. All nodes use the instance type specified in the command.

> **Note**
>
> If you have not previously created the default EMR service role and EC2 instance profile, type `aws emr create-default-roles` to create them before typing the `create-cluster` subcommand.

For more information about using Amazon EMR commands in the AWS CLI, see https://docs.aws.amazon.com/cli/latest/reference/emr.

View Ganglia metrics

> **Note**
>
> The last release of Amazon EMR to include Ganglia was Amazon EMR 6.15.0. To monitor your cluster, releases higher than 6.15.0 include the Amazon CloudWatch agent.

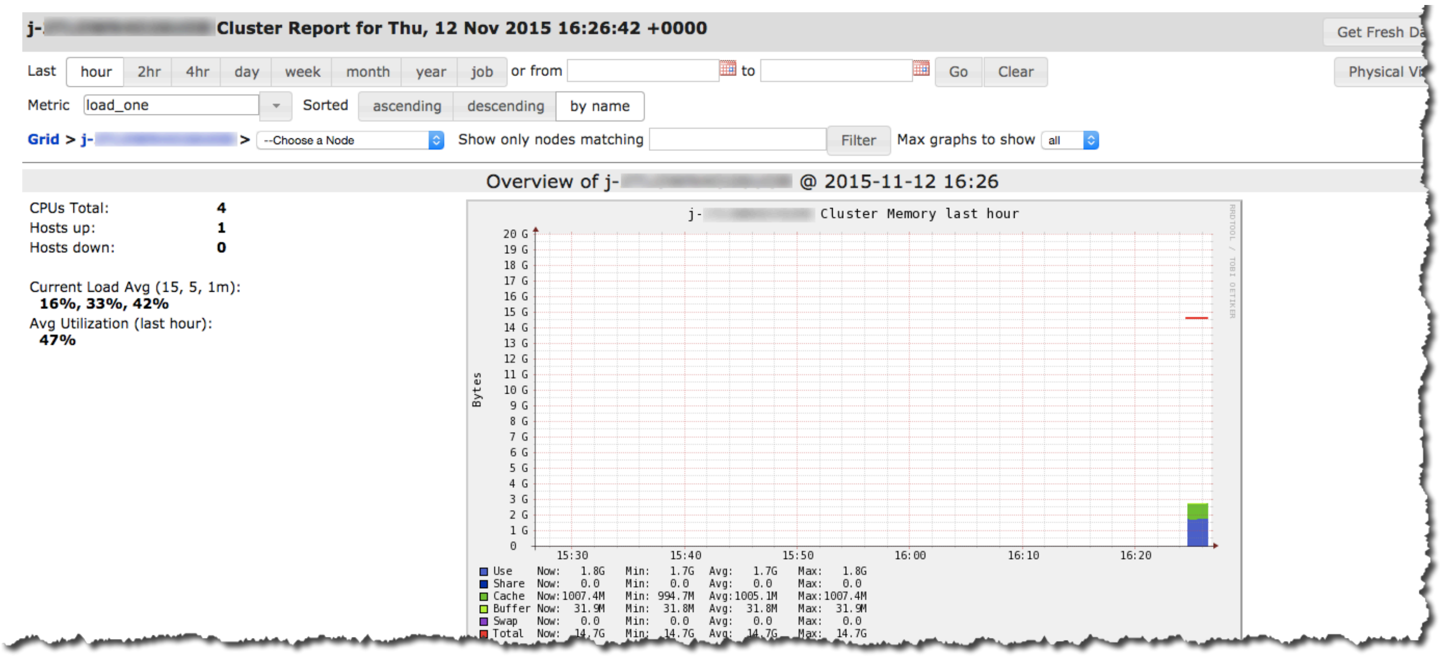
Ganglia provides a web-based user interface that you can use to view the metrics Ganglia collects. When you run Ganglia on Amazon EMR, the web interface runs on the master node and can be viewed using port forwarding, also known as creating an SSH tunnel. For more information about viewing web interfaces on Amazon EMR, see View web interfaces hosted on EMR clusters in the *Amazon EMR Management Guide*.

To view the Ganglia web interface

1. Use SSH to tunnel into the master node and create a secure connection. For information about how to create an SSH tunnel to the master node, see Option 2, part 1: Set up an SSH tunnel to the master node using dynamic port forwarding in the *Amazon EMR Management Guide*.
2. Install a web browser with a proxy tool, such as the FoxyProxy plug-in for Firefox, to create a SOCKS proxy for domains of the type *ec2*.amazonaws.com*. For more information, see

[Option 2, part 2: Configure proxy settings to view websites hosted on the master node](#) in the *Amazon EMR Management Guide*.

3. With the proxy set and the SSH connection open, you can view the Ganglia UI by opening a browser window with http://*master-public-dns-name*/ganglia/, where *master-public-dns-name* is the public DNS address of the master server in the EMR cluster.



Hadoop and Spark metrics in Ganglia

> **Note**
>
> The last release of Amazon EMR to include Ganglia was Amazon EMR 6.15.0. To monitor your cluster, releases higher than 6.15.0 include the [Amazon CloudWatch agent](#).

Ganglia reports Hadoop metrics for each instance. The various types of metrics are prefixed by category: distributed file system (dfs.*), Java virtual machine (jvm.*), MapReduce (mapred.*), and remote procedure calls (rpc.*).

YARN-based Ganglia metrics such as Spark and Hadoop are not available for EMR release versions 4.4.0 and 4.5.0. Use a later version to use these metrics.

Ganglia metrics for Spark generally have prefixes for YARN application ID and Spark DAGScheduler. So prefixes follow this form:

- DAGScheduler.*
- application_xxxxxxxxxx_xxxx.driver.*
- application_xxxxxxxxxx_xxxx.executor.*

Ganglia release history

> **Note**
>
> The last release of Amazon EMR to include Ganglia was Amazon EMR 6.15.0. To monitor your cluster, releases higher than 6.15.0 include the Amazon CloudWatch agent.

The following table lists the version of Ganglia included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in Amazon EMR 7.x release versions, Amazon EMR 6.x release versions, or Amazon EMR 5.x release versions.

Ganglia version information

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.36.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-6.15.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.14.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-6.13.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.12.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-6.11.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.11.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-6.10.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.10.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-6.9.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.9.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-6.8.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.8.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-6.7.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.36.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.36.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.6.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.35.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.5.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-6.4.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.3.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-6.3.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.2.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-6.2.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.1.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-6.1.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.0.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-6.0.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.34.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.33.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.33.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.32.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.32.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.31.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.31.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.30.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.30.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.30.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.29.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.28.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.28.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.27.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.27.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.26.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.25.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.24.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.24.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.23.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.23.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.22.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.21.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.21.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.21.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.20.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.20.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.19.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.19.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.18.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.18.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.17.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.17.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.17.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.16.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.16.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.15.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.15.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.14.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.14.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.14.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.13.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.13.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.12.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.12.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.12.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.12.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.11.4	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.11.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.11.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.11.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.11.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.10.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.10.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.9.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.9.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.8.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.8.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.8.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.8.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.7.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.7.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.6.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.6.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.5.4	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.5.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.5.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.5.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.5.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.4.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.4.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.3.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.3.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.3.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.2.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.2.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.2.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.2.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.1.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.1.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.0.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-5.0.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.0.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.0.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-4.9.6	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.9.5	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.9.4	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-4.9.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.9.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.9.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-4.8.5	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.8.4	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.8.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-4.8.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.8.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.8.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-4.7.4	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.7.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.7.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-4.7.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.7.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.6.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-4.6.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.5.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.4.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon EMR Release label	Ganglia Version	Components installed with Ganglia
emr-4.3.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.2.0	3.6.0	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Apache Hadoop

Apache Hadoop is an open-source Java software framework that supports massive data processing across a cluster of instances. It can run on a single instance or thousands of instances. Hadoop uses various processing models, such as MapReduce and Tez, to distribute processing across multiple instances and also uses a distributed file system called HDFS to store data across multiple instances. Hadoop monitors the health of instances in the cluster and can recover from the failure of one or more nodes. In this way, Hadoop provides increased processing and storage capacity, as well as high availability. For more information, see theHadoop documentation.

The following table lists the version of Hadoop included in the latest release of the Amazon EMR 7.x series, along with the components that Amazon EMR installs with Hadoop.

For the version of components installed with Hadoop in this release, see Release 7.2.0 Component Versions.

Hadoop version information for emr-7.2.0

Amazon EMR Release Label	Hadoop Version	Components Installed With Hadoop
emr-7.2.0	Hadoop 3.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

The following table lists the version of Hadoop included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with Hadoop.

For the version of components installed with Hadoop in this release, see Release 6.15.0 Component Versions.

Hadoop version information for emr-6.15.0

Amazon EMR Release Label	Hadoop Version	Components Installed With Hadoop
emr-6.15.0	Hadoop 3.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

The following table lists the version of Hadoop included in the latest release of the Amazon EMR 5.x series, along with the components that Amazon EMR installs with Hadoop.

For the version of components installed with Hadoop in this release, see Release 5.36.2 Component Versions.

Hadoop version information for emr-5.36.2

Amazon EMR Release Label	Hadoop Version	Components Installed With Hadoop
emr-5.36.2	Hadoop 2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-ya

Amazon EMR Release Label	Hadoop Version	Components Installed With Hadoop
		rn-resourcemanager, hadoop-yarn-timeline-server

Beginning with Amazon EMR 5.18.0, you can use the Amazon EMR artifact repository to build your job code against the exact versions of libraries and dependencies that are available with specific Amazon EMR releases. For more information, see Checking dependencies using the Amazon EMR artifact repository.

Topics

- Configure Hadoop
- Transparent encryption in HDFS on Amazon EMR
- Create or run a Hadoop application
- Read restored objects
- Turn on non-uniform memory access awareness for YARN containers
- Hadoop version history

Configure Hadoop

The following sections give default configuration settings for Hadoop daemons, tasks, and HDFS.

Topics

- Task configuration
- Hadoop daemon configuration settings
- HDFS configuration

Task configuration

You can set configuration variables to tune the performance of your MapReduce jobs. This section provides the default values for important settings. Default values vary based on the EC2 instance type of the node used in the cluster. HBase is available when using Amazon EMR release version

4.6.0 and later. Different defaults are used when HBase is installed. Those values are provided along with the initial defaults.

Hadoop 2 uses two parameters, `mapreduce.map.java.opts` and `mapreduce.reduce.java.opts`, to configure memory for map and reduce JVMs respectively. These replace the single `mapreduce.map.java.opts` configuration option from earlier Hadoop versions.

Similarly, `mapred.job.jvm.num.tasks` replaces `mapred.job.reuse.jvm.num.tasks` in Hadoop 2.7.2 and later. Amazon EMR sets this value to 20 regardless of EC2 instance type. You can override this setting using the `mapred-site` configuration classification. Setting a value of `-1` indicates that a JVM can be re-used for an infinite number of tasks within a single job, and a value of 1 indicates that a new JVM is spawned for each task.

For example, to set the value of `mapred.job.jvm.num.tasks` to -1 you can create a file with the following contents:

```
[
    {
      "Classification": "mapred-site",
      "Properties": {
        "mapred.job.jvm.num.tasks": "-1"
      }
    }
  ]
```

When you use the `create-cluster` command or `modify-instance-groups` command from the AWS CLI, you can then reference the JSON configuration file. In the following example, the configuration file is saved as `myConfig.json` and stored in Amazon S3.

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge \
--instance-count 3 --applications Name=Hadoop --configurations https://
s3.amazonaws.com/mybucket/myfolder/myConfig.json \
```

```
--use-default-roles
```

You can change default values listed below using the `mapred-site` configuration classification in the same way, and set multiple values and multiple configuration classifications using a single JSON file. For more information, see Configure applications.

With Amazon EMR version 5.21.0 and later, you can override cluster configurations and specify additional configuration classifications for each instance group in a running cluster. You do this by using the Amazon EMR console, the AWS Command Line Interface (AWS CLI), or the AWS SDK. For more information, see Supplying a Configuration for an Instance Group in a Running Cluster.

Default values for task configuration settings

Instance Types

- c1 instances
- c3 instances
- c4 instances
- c5 instances
- c5a instances
- c5ad instances
- c5d instances
- c5n instances
- c6a instances
- c6g instances
- c6gd instances
- c6gn instances
- c6i instances
- c6id instances
- c6in instances
- c7a instances
- c7g instances

- c7gd instances
- c7gn instances
- c7i instances
- d2 instances
- d3 instances
- d3en instances
- g3 instances
- g3s instances
- g4dn instances
- g5 instances
- h1 instances
- i2 instances
- i3 instances
- i3en instances
- i4g instances
- i4i instances
- im4gn instances
- is4gen instances
- m1 instances
- m2 instances
- m3 instances
- m4 instances
- m5 instances
- m5a instances
- m5ad instances
- m5d instances
- m5dn instances
- m5n instances

- m5zn instances
- m6a instances
- m6g instances
- m6gd instances
- m6i instances
- m6id instances
- m6idn instances
- m6in instances
- m7a instances
- m7g instances
- m7gd instances
- m7i instances
- m7i-flex instances
- p2 instances
- p3 instances
- p5 instances
- r3 instances
- r4 instances
- r5 instances
- r5a instances
- r5ad instances
- r5b instances
- r5d instances
- r5dn instances
- r5n instances
- r6a instances
- r6g instances
- r6gd instances

c1 instances

c1.medium

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx288m	-Xmx288m
mapreduce.java.opts	-Xmx288m	-Xmx288m
mapreduce.map.memory.mb	512	512
mapreduce.reduce.memory.mb	512	512
yarn.app.mapreduce.am.resource.mb	512	512
yarn.scheduler.minimum-allocation-mb	256	256

Configuration option	Default value	With HBase installed
yarn.scheduler.maximum-allocation-mb	512	512
yarn.nodemanager.resource.memory-mb	1024	512

c1.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx864m	-Xmx864m
mapreduce.java.opts	-Xmx1536m	-Xmx1536m
mapreduce.map.memory.mb	1024	1024
mapreduce.reduce.memory.mb	2048	2048
yarn.app.mapreduce.am.resource.mb	2048	2048
yarn.scheduler.minimum-allocation-mb	256	256
yarn.scheduler.maximum-allocation-mb	2048	2560
yarn.nodemanager.resource.memory-mb	5120	2560

c3 instances

c3.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m

Configuration option	Default value	With HBase installed
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

c3.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1152m	-Xmx1152m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1440	1440
mapreduce.reduce.memory.mb	2880	2880
yarn.app.mapreduce.am.resource.mb	2880	2880
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11520	5760

Configuration option	Default value	With HBase installed
yarn.nodemanager.resource.memory-mb	11520	5760

c3.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1152m	-Xmx1152m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1440	1440
mapreduce.reduce.memory.mb	2880	2880
yarn.app.mapreduce.am.resource.mb	2880	2880
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23040	11520
yarn.nodemanager.resource.memory-mb	23040	11520

c3.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1331m	-Xmx1331m
mapreduce.java.opts	-Xmx2662m	-Xmx2662m
mapreduce.map.memory.mb	1664	1664

Configuration option	Default value	With HBase installed
mapreduce.reduce.memory.mb	3328	3328
yarn.app.mapreduce.am.resource.mb	3328	3328
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	53248	26624
yarn.nodemanager.resource.memory-mb	53248	26624

c4 instances

c4.large

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx717m	-Xmx717m
mapreduce.java.opts	-Xmx1434m	-Xmx1434m
mapreduce.map.memory.mb	896	896
mapreduce.reduce.memory.mb	1792	1792
yarn.app.mapreduce.am.resource.mb	1792	1792
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1792	896

Configuration option	Default value	With HBase installed
yarn.nodemanager.resource.memory-mb	1792	896

c4.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

c4.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1152m	-Xmx1152m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1440	1440

Configuration option	Default value	With HBase installed
mapreduce.reduce.memory.mb	2880	2880
yarn.app.mapreduce.am.resource.mb	2880	2880
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11520	5760
yarn.nodemanager.resource.memory-mb	11520	5760

c4.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1152m	-Xmx1152m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1440	1440
mapreduce.reduce.memory.mb	2880	2880
yarn.app.mapreduce.am.resource.mb	2880	2880
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23040	11520
yarn.nodemanager.resource.memory-mb	23040	11520

c4.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1183m	-Xmx1183m
mapreduce.java.opts	-Xmx2366m	-Xmx2366m
mapreduce.map.memory.mb	1479	1479
mapreduce.reduce.memory.mb	2958	2958
yarn.app.mapreduce.am.resource.mb	2958	2958
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	53248	26624
yarn.nodemanager.resource.memory-mb	53248	26624

c5 instances

c5.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072
yarn.app.mapreduce.am.resource.mb	3072	3072

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	6144	3072
yarn.nodemanager.resource.memory-mb	6144	3072

c5.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072
yarn.app.mapreduce.am.resource.mb	3072	3072
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	12288	6144
yarn.nodemanager.resource.memory-mb	12288	6144

c5.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072
yarn.app.mapreduce.am.resource.mb	3072	3072
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

c5.9xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1456m	-Xmx1456m
mapreduce.java.opts	-Xmx2912m	-Xmx2912m
mapreduce.map.memory.mb	1820	1820
mapreduce.reduce.memory.mb	3640	3640
yarn.app.mapreduce.am.resource.mb	3640	3640

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	65536	32768
yarn.nodemanager.resource.memory-mb	65536	32768

c5.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1502m	-Xmx1502m
mapreduce.java.opts	-Xmx3004m	-Xmx3004m
mapreduce.map.memory.mb	1877	1877
mapreduce.reduce.memory.mb	3754	3754
yarn.app.mapreduce.am.resource.mb	3754	3754
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	90112	30048
yarn.nodemanager.resource.memory-mb	90112	30048

c5.18xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1547m	-Xmx1547m
mapreduce.java.opts	-Xmx3094m	-Xmx3094m
mapreduce.map.memory.mb	1934	1934
mapreduce.reduce.memory.mb	3868	3868
yarn.app.mapreduce.am.resource.mb	3868	3868
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	139264	30960
yarn.nodemanager.resource.memory-mb	139264	30960

c5.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1570m	-Xmx1570m
mapreduce.java.opts	-Xmx3140m	-Xmx3140m
mapreduce.map.memory.mb	1963	1963
mapreduce.reduce.memory.mb	3926	3926
yarn.app.mapreduce.am.resource.mb	3926	3926

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31376
yarn.nodemanager.resource.memory-mb	188416	31376

c5a instances

c5a.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

c5a.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

c5a.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

c5a.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

c5a.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1502m	-Xmx1502m
mapreduce.java.opts	-Xmx3004m	-Xmx3004m
mapreduce.map.memory.mb	1877	1877
mapreduce.reduce.memory.mb	3754	3754
yarn.app.mapreduce.am.resource.mb	3754	3754
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	90112	30048
yarn.nodemanager.resource.memory-mb	90112	30048

c5a.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

c5a.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29840
yarn.nodemanager.resource.memory-mb	179200	29840

c5ad instances

c5ad.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

c5ad.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

c5ad.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

c5ad.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

c5ad.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

c5ad.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

c5ad.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29840
yarn.nodemanager.resource.memory-mb	179200	29840

c5d instances

c5d.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072
yarn.app.mapreduce.am.resource.mb	3072	3072

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	6144	3072
yarn.nodemanager.resource.memory-mb	6144	3072

c5d.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072
yarn.app.mapreduce.am.resource.mb	3072	3072
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	12288	6144
yarn.nodemanager.resource.memory-mb	12288	6144

c5d.4xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx1229m	-Xmx1229m
`mapreduce.java.opts`	-Xmx2458m	-Xmx2458m
`mapreduce.map.memory.mb`	1536	1536
`mapreduce.reduce.memory.mb`	3072	3072
`yarn.app.mapreduce.am.resource.mb`	3072	3072
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	24576	12288
`yarn.nodemanager.resource.memory-mb`	24576	12288

c5d.9xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx1456m	-Xmx1456m
`mapreduce.java.opts`	-Xmx2912m	-Xmx2912m
`mapreduce.map.memory.mb`	1820	1820
`mapreduce.reduce.memory.mb`	3640	3640
`yarn.app.mapreduce.am.resource.mb`	3640	3640

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	65536	32768
yarn.nodemanager.resource.memory-mb	65536	32768

c5d.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1502m	-Xmx1502m
mapreduce.java.opts	-Xmx3004m	-Xmx3004m
mapreduce.map.memory.mb	1877	1877
mapreduce.reduce.memory.mb	3754	3754
yarn.app.mapreduce.am.resource.mb	3754	3754
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	90112	30048
yarn.nodemanager.resource.memory-mb	90112	30048

c5d.18xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1547m	-Xmx1547m
mapreduce.java.opts	-Xmx3094m	-Xmx3094m
mapreduce.map.memory.mb	1934	1934
mapreduce.reduce.memory.mb	3868	3868
yarn.app.mapreduce.am.resource.mb	3868	3868
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	139264	30960
yarn.nodemanager.resource.memory-mb	139264	30960

c5d.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1570m	-Xmx1570m
mapreduce.java.opts	-Xmx3140m	-Xmx3140m
mapreduce.map.memory.mb	1963	1963
mapreduce.reduce.memory.mb	3926	3926
yarn.app.mapreduce.am.resource.mb	3926	3926

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31376
yarn.nodemanager.resource.memory-mb	188416	31376

c5n instances

c5n.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1613m	-Xmx1613m
mapreduce.java.opts	-Xmx3226m	-Xmx3226m
mapreduce.map.memory.mb	2016	2016
mapreduce.reduce.memory.mb	4032	4032
yarn.app.mapreduce.am.resource.mb	4032	4032
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	8064	4032
yarn.nodemanager.resource.memory-mb	8064	4032

c5n.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1613m	-Xmx1613m
mapreduce.java.opts	-Xmx3226m	-Xmx3226m
mapreduce.map.memory.mb	2016	2016
mapreduce.reduce.memory.mb	4032	4032
yarn.app.mapreduce.am.resource.mb	4032	4032
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	16128	8064
yarn.nodemanager.resource.memory-mb	16128	8064

c5n.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1741m	-Xmx1741m
mapreduce.java.opts	-Xmx3482m	-Xmx3482m
mapreduce.map.memory.mb	2176	2176
mapreduce.reduce.memory.mb	4352	4352
yarn.app.mapreduce.am.resource.mb	4352	4352

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	34816	17408
yarn.nodemanager.resource.memory-mb	34816	17408

c5n.9xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2002m	-Xmx2002m
mapreduce.java.opts	-Xmx4004m	-Xmx4004m
mapreduce.map.memory.mb	2503	2503
mapreduce.reduce.memory.mb	5006	5006
yarn.app.mapreduce.am.resource.mb	5006	5006
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	90112	30040
yarn.nodemanager.resource.memory-mb	90112	30040

c5n.18xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2094m	-Xmx2094m
mapreduce.java.opts	-Xmx4188m	-Xmx4188m
mapreduce.map.memory.mb	2617	2617
mapreduce.reduce.memory.mb	5234	5234
yarn.app.mapreduce.am.resource.mb	5234	5234
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31396
yarn.nodemanager.resource.memory-mb	188416	31396

c6a instances

c6a.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

c6a.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

c6a.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

c6a.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

c6a.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

c6a.16xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx1459m	-Xmx1459m
`mapreduce.java.opts`	-Xmx2918m	-Xmx2918m
`mapreduce.map.memory.mb`	1824	1824
`mapreduce.reduce.memory.mb`	3648	3648
`yarn.app.mapreduce.am.resource.mb`	3648	3648
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	116736	29184
`yarn.nodemanager.resource.memory-mb`	116736	29184

c6a.24xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx1494m	-Xmx1494m
`mapreduce.java.opts`	-Xmx2988m	-Xmx2988m
`mapreduce.map.memory.mb`	1867	1867
`mapreduce.reduce.memory.mb`	3734	3734
`yarn.app.mapreduce.am.resource.mb`	3734	3734

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29840
yarn.nodemanager.resource.memory-mb	179200	29840

c6a.32xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1510m	-Xmx1510m
mapreduce.java.opts	-Xmx3020m	-Xmx3020m
mapreduce.map.memory.mb	1888	1888
mapreduce.reduce.memory.mb	3776	3776
yarn.app.mapreduce.am.resource.mb	3776	3776
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

c6a.48xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1527m	-Xmx1527m
mapreduce.java.opts	-Xmx3054m	-Xmx3054m
mapreduce.map.memory.mb	1909	1909
mapreduce.reduce.memory.mb	3818	3818
yarn.app.mapreduce.am.resource.mb	3818	3818
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30608
yarn.nodemanager.resource.memory-mb	366592	30608

c6g instances

c6g.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

c6g.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

c6g.4xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx1171m	-Xmx1171m
`mapreduce.java.opts`	-Xmx2342m	-Xmx2342m
`mapreduce.map.memory.mb`	1464	1464
`mapreduce.reduce.memory.mb`	2928	2928
`yarn.app.mapreduce.am.resource.mb`	2928	2928
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	23424	11712
`yarn.nodemanager.resource.memory-mb`	23424	11712

c6g.8xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx1357m	-Xmx1357m
`mapreduce.java.opts`	-Xmx2714m	-Xmx2714m
`mapreduce.map.memory.mb`	1696	1696
`mapreduce.reduce.memory.mb`	3392	3392
`yarn.app.mapreduce.am.resource.mb`	3392	3392

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

c6g.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

c6g.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

c6gd instances

c6gd.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

c6gd.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

c6gd.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

c6gd.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

c6gd.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

c6gd.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

c6gn instances

c6gn.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

c6gn.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

c6gn.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

c6gn.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

c6gn.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

c6gn.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

c6i instances

c6i.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

c6i.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

c6i.4xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx1171m	-Xmx1171m
`mapreduce.java.opts`	-Xmx2342m	-Xmx2342m
`mapreduce.map.memory.mb`	1464	1464
`mapreduce.reduce.memory.mb`	2928	2928
`yarn.app.mapreduce.am.resource.mb`	2928	2928
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	23424	11712
`yarn.nodemanager.resource.memory-mb`	23424	11712

c6i.8xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx1357m	-Xmx1357m
`mapreduce.java.opts`	-Xmx2714m	-Xmx2714m
`mapreduce.map.memory.mb`	1696	1696
`mapreduce.reduce.memory.mb`	3392	3392
`yarn.app.mapreduce.am.resource.mb`	3392	3392

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

c6i.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

c6i.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

c6i.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29840
yarn.nodemanager.resource.memory-mb	179200	29840

c6i.32xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1510m	-Xmx1510m
mapreduce.java.opts	-Xmx3020m	-Xmx3020m
mapreduce.map.memory.mb	1888	1888
mapreduce.reduce.memory.mb	3776	3776
yarn.app.mapreduce.am.resource.mb	3776	3776
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

c6id instances

c6id.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

c6id.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

c6id.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

c6id.8xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx1357m	-Xmx1357m
`mapreduce.java.opts`	-Xmx2714m	-Xmx2714m
`mapreduce.map.memory.mb`	1696	1696
`mapreduce.reduce.memory.mb`	3392	3392
`yarn.app.mapreduce.am.resource.mb`	3392	3392
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	54272	27136
`yarn.nodemanager.resource.memory-mb`	54272	27136

c6id.12xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx1425m	-Xmx1425m
`mapreduce.java.opts`	-Xmx2850m	-Xmx2850m
`mapreduce.map.memory.mb`	1781	1781
`mapreduce.reduce.memory.mb`	3562	3562
`yarn.app.mapreduce.am.resource.mb`	3562	3562

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

c6id.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

c6id.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29840
yarn.nodemanager.resource.memory-mb	179200	29840

c6id.32xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1510m	-Xmx1510m
mapreduce.java.opts	-Xmx3020m	-Xmx3020m
mapreduce.map.memory.mb	1888	1888
mapreduce.reduce.memory.mb	3776	3776
yarn.app.mapreduce.am.resource.mb	3776	3776

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

c6in instances

c6in.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

c6in.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

c6in.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

c6in.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

c6in.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

c6in.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

c6in.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29840
yarn.nodemanager.resource.memory-mb	179200	29840

c6in.32xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1510m	-Xmx1510m
mapreduce.java.opts	-Xmx3020m	-Xmx3020m
mapreduce.map.memory.mb	1888	1888
mapreduce.reduce.memory.mb	3776	3776
yarn.app.mapreduce.am.resource.mb	3776	3776
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

c7a instances

c7a.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

c7a.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

c7a.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

c7a.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

c7a.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

c7a.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

c7a.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29840
yarn.nodemanager.resource.memory-mb	179200	29840

c7a.32xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1510m	-Xmx1510m
mapreduce.java.opts	-Xmx3020m	-Xmx3020m
mapreduce.map.memory.mb	1888	1888
mapreduce.reduce.memory.mb	3776	3776
yarn.app.mapreduce.am.resource.mb	3776	3776
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

c7a.48xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1527m	-Xmx1527m
mapreduce.java.opts	-Xmx3054m	-Xmx3054m
mapreduce.map.memory.mb	1909	1909
mapreduce.reduce.memory.mb	3818	3818
yarn.app.mapreduce.am.resource.mb	3818	3818
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30608
yarn.nodemanager.resource.memory-mb	366592	30608

c7g instances

c7g.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

Configuration option	Default value	With HBase installed
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	5632	2816
`yarn.nodemanager.resource.memory-mb`	5632	2816

c7g.2xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx1171m	-Xmx1171m
`mapreduce.java.opts`	-Xmx2342m	-Xmx2342m
`mapreduce.map.memory.mb`	1464	1464
`mapreduce.reduce.memory.mb`	2928	2928
`yarn.app.mapreduce.am.resource.mb`	2928	2928
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	11712	5856
`yarn.nodemanager.resource.memory-mb`	11712	5856

c7g.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

c7g.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

c7g.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

c7g.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

c7gd instances

c7gd.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

c7gd.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

c7gd.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

c7gd.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

c7gd.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

c7gd.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

c7gn instances

c7gn.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

c7gn.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

c7gn.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

c7gn.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

c7gn.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

c7gn.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

c7i instances

c7i.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

c7i.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

c7i.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

c7i.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

c7i.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

c7i.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

c7i.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29840
yarn.nodemanager.resource.memory-mb	179200	29840

c7i.48xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1527m	-Xmx1527m
mapreduce.java.opts	-Xmx3054m	-Xmx3054m
mapreduce.map.memory.mb	1909	1909
mapreduce.reduce.memory.mb	3818	3818
yarn.app.mapreduce.am.resource.mb	3818	3818
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30608
yarn.nodemanager.resource.memory-mb	366592	30608

d2 instances

d2.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

d2.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

d2.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

d2.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2417m	-Xmx2417m
mapreduce.java.opts	-Xmx4834m	-Xmx4834m
mapreduce.map.memory.mb	3021	3021
mapreduce.reduce.memory.mb	6042	6042
yarn.app.mapreduce.am.resource.mb	6042	6042
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30194
yarn.nodemanager.resource.memory-mb	241664	30194

d3 instances

d3.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

d3.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

d3.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

d3.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

d3en instances

d3en.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

d3en.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

d3en.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

d3en.6xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.java.opts	-Xmx5700m	-Xmx5700m
mapreduce.map.memory.mb	3563	3563
mapreduce.reduce.memory.mb	7126	7126
yarn.app.mapreduce.am.resource.mb	7126	7126
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	28496
yarn.nodemanager.resource.memory-mb	85504	28496

d3en.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

d3en.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

g3 instances

g3.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

g3.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

g3.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

g3s instances

g3s.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

g4dn instances

g4dn.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144
yarn.app.mapreduce.am.resource.mb	6144	6144
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	12288	6144
yarn.nodemanager.resource.memory-mb	12288	6144

g4dn.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144
yarn.app.mapreduce.am.resource.mb	6144	6144

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

g4dn.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2867m	-Xmx2867m
mapreduce.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.map.memory.mb	3584	3584
mapreduce.reduce.memory.mb	7168	7168
yarn.app.mapreduce.am.resource.mb	7168	7168
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

g4dn.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	122880	30720
yarn.nodemanager.resource.memory-mb	122880	30720

g4dn.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3140m	-Xmx3140m
mapreduce.java.opts	-Xmx6280m	-Xmx6280m
mapreduce.map.memory.mb	3925	3925
mapreduce.reduce.memory.mb	7850	7850
yarn.app.mapreduce.am.resource.mb	7850	7850

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31416
yarn.nodemanager.resource.memory-mb	188416	31416

g4dn.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3174m	-Xmx3174m
mapreduce.java.opts	-Xmx6348m	-Xmx6348m
mapreduce.map.memory.mb	3968	3968
mapreduce.reduce.memory.mb	7936	7936
yarn.app.mapreduce.am.resource.mb	7936	7936
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744

g5 instances

g5.xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx2342m	-Xmx2342m
`mapreduce.java.opts`	-Xmx4684m	-Xmx4684m
`mapreduce.map.memory.mb`	2928	2928
`mapreduce.reduce.memory.mb`	5856	5856
`yarn.app.mapreduce.am.resource.mb`	5856	5856
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	11712	5856
`yarn.nodemanager.resource.memory-mb`	11712	5856

g5.2xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx2342m	-Xmx2342m
`mapreduce.java.opts`	-Xmx4684m	-Xmx4684m
`mapreduce.map.memory.mb`	2928	2928
`mapreduce.reduce.memory.mb`	5856	5856
`yarn.app.mapreduce.am.resource.mb`	5856	5856

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

g5.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

g5.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

g5.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

g5.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

g5.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

g5.48xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3089m	-Xmx3089m
mapreduce.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.map.memory.mb	3861	3861
mapreduce.reduce.memory.mb	7722	7722
yarn.app.mapreduce.am.resource.mb	7722	7722

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30952
yarn.nodemanager.resource.memory-mb	741376	30952

h1 instances

h1.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144
yarn.app.mapreduce.am.resource.mb	6144	6144
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

h1.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2867m	-Xmx2867m
mapreduce.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.map.memory.mb	3584	3584
mapreduce.reduce.memory.mb	7168	7168
yarn.app.mapreduce.am.resource.mb	7168	7168
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

h1.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	122880	30720
yarn.nodemanager.resource.memory-mb	122880	30720

h1.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3174m	-Xmx3174m
mapreduce.java.opts	-Xmx6348m	-Xmx6348m
mapreduce.map.memory.mb	3968	3968
mapreduce.reduce.memory.mb	7936	7936
yarn.app.mapreduce.am.resource.mb	7936	7936
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744

i2 instances

i2.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

i2.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

i2.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

i2.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

i3 instances

i3.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

i3.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

i3.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

i3.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

i3.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

i3en instances

i3en.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4915m	-Xmx4915m
mapreduce.java.opts	-Xmx9830m	-Xmx9830m
mapreduce.map.memory.mb	6144	6144
mapreduce.reduce.memory.mb	12288	12288
yarn.app.mapreduce.am.resource.mb	12288	12288
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

i3en.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.java.opts	-Xmx11468m	-Xmx11468m
mapreduce.map.memory.mb	7168	7168
mapreduce.reduce.memory.mb	14336	14336
yarn.app.mapreduce.am.resource.mb	14336	14336

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

i3en.3xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6007m	-Xmx6007m
mapreduce.java.opts	-Xmx12014m	-Xmx12014m
mapreduce.map.memory.mb	7509	7509
mapreduce.reduce.memory.mb	15018	15018
yarn.app.mapreduce.am.resource.mb	15018	15018
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	90112	30040
yarn.nodemanager.resource.memory-mb	90112	30040

i3en.6xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6281m	-Xmx6281m
mapreduce.java.opts	-Xmx12562m	-Xmx12562m
mapreduce.map.memory.mb	7851	7851
mapreduce.reduce.memory.mb	15702	15702
yarn.app.mapreduce.am.resource.mb	15702	15702
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31396
yarn.nodemanager.resource.memory-mb	188416	31396

i3en.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6417m	-Xmx6417m
mapreduce.java.opts	-Xmx12834m	-Xmx12834m
mapreduce.map.memory.mb	8021	8021
mapreduce.reduce.memory.mb	16042	16042
yarn.app.mapreduce.am.resource.mb	16042	16042

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32100
yarn.nodemanager.resource.memory-mb	385024	32100

i3en.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6486m	-Xmx6486m
mapreduce.java.opts	-Xmx12972m	-Xmx12972m
mapreduce.map.memory.mb	8107	8107
mapreduce.reduce.memory.mb	16214	16214
yarn.app.mapreduce.am.resource.mb	16214	16214
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	778240	32396
yarn.nodemanager.resource.memory-mb	778240	32396

i4g instances

i4g.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

i4g.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

i4g.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

i4g.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

i4g.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

i4i instances

i4i.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

i4i.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

i4i.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

i4i.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

i4i.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

i4i.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

i4i.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

i4i.32xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	7744	7744
mapreduce.reduce.memory.mb	15488	15488
yarn.app.mapreduce.am.resource.mb	15488	15488
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

im4gn instances

im4gn.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

im4gn.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

im4gn.4xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx2714m	-Xmx2714m
`mapreduce.java.opts`	-Xmx5428m	-Xmx5428m
`mapreduce.map.memory.mb`	3392	3392
`mapreduce.reduce.memory.mb`	6784	6784
`yarn.app.mapreduce.am.resource.mb`	6784	6784
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	54272	27136
`yarn.nodemanager.resource.memory-mb`	54272	27136

im4gn.8xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx2918m	-Xmx2918m
`mapreduce.java.opts`	-Xmx5836m	-Xmx5836m
`mapreduce.map.memory.mb`	3648	3648
`mapreduce.reduce.memory.mb`	7296	7296
`yarn.app.mapreduce.am.resource.mb`	7296	7296

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

im4gn.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

is4gen instances

is4gen.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3514m	-Xmx3514m
mapreduce.java.opts	-Xmx7028m	-Xmx7028m
mapreduce.map.memory.mb	4393	4393
mapreduce.reduce.memory.mb	8786	8786
yarn.app.mapreduce.am.resource.mb	8786	8786
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	17572.12	8786.06
yarn.nodemanager.resource.memory-mb	17572.12	8786.06

is4gen.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3866m	-Xmx3866m
mapreduce.java.opts	-Xmx7732m	-Xmx7732m
mapreduce.map.memory.mb	4832	4832
mapreduce.reduce.memory.mb	9664	9664
yarn.app.mapreduce.am.resource.mb	9664	9664

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	38656	19328
yarn.nodemanager.resource.memory-mb	38656	19328

is4gen.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4275m	-Xmx4275m
mapreduce.java.opts	-Xmx8550m	-Xmx8550m
mapreduce.map.memory.mb	5344	5344
mapreduce.reduce.memory.mb	10688	10688
yarn.app.mapreduce.am.resource.mb	10688	10688
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32064
yarn.nodemanager.resource.memory-mb	85504	32064

is4gen.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4480m	-Xmx4480m
mapreduce.java.opts	-Xmx8960m	-Xmx8960m
mapreduce.map.memory.mb	5600	5600
mapreduce.reduce.memory.mb	11200	11200
yarn.app.mapreduce.am.resource.mb	11200	11200
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	22400
yarn.nodemanager.resource.memory-mb	179200	22400

m1 instances

m1.small

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx288m	-Xmx288m
mapreduce.java.opts	-Xmx288m	-Xmx288m
mapreduce.map.memory.mb	512	512
mapreduce.reduce.memory.mb	512	512
yarn.app.mapreduce.am.resource.mb	512	512

Configuration option	Default value	With HBase installed
`yarn.scheduler.minimum-allocation-mb`	256	256
`yarn.scheduler.maximum-allocation-mb`	512	512
`yarn.nodemanager.resource.memory-mb`	1024	512

m1.medium

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx512m	-Xmx512m
`mapreduce.java.opts`	-Xmx768m	-Xmx768m
`mapreduce.map.memory.mb`	768	768
`mapreduce.reduce.memory.mb`	1024	1024
`yarn.app.mapreduce.am.resource.mb`	1024	1024
`yarn.scheduler.minimum-allocation-mb`	256	256
`yarn.scheduler.maximum-allocation-mb`	2048	1024
`yarn.nodemanager.resource.memory-mb`	2048	1024

m1.large

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx512m	-Xmx512m
`mapreduce.java.opts`	-Xmx1024m	-Xmx1024m
`mapreduce.map.memory.mb`	768	768
`mapreduce.reduce.memory.mb`	1536	1536
`yarn.app.mapreduce.am.resource.mb`	1536	1536
`yarn.scheduler.minimum-allocation-mb`	256	256
`yarn.scheduler.maximum-allocation-mb`	3072	2560
`yarn.nodemanager.resource.memory-mb`	5120	2560

m1.xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx512m	-Xmx512m
`mapreduce.java.opts`	-Xmx1536m	-Xmx1536m
`mapreduce.map.memory.mb`	768	768
`mapreduce.reduce.memory.mb`	2048	2048
`yarn.app.mapreduce.am.resource.mb`	2048	2048

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	256	256
yarn.scheduler.maximum-allocation-mb	8192	6144
yarn.nodemanager.resource.memory-mb	12288	6144

m2 instances

m2.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx864m	-Xmx864m
mapreduce.java.opts	-Xmx1536m	-Xmx1536m
mapreduce.map.memory.mb	1024	1024
mapreduce.reduce.memory.mb	2048	2048
yarn.app.mapreduce.am.resource.mb	2048	2048
yarn.scheduler.minimum-allocation-mb	256	256
yarn.scheduler.maximum-allocation-mb	7168	7168
yarn.nodemanager.resource.memory-mb	14336	7168

m2.2xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx1280m	-Xmx1280m
`mapreduce.java.opts`	-Xmx2304m	-Xmx2304m
`mapreduce.map.memory.mb`	1536	1536
`mapreduce.reduce.memory.mb`	2560	2560
`yarn.app.mapreduce.am.resource.mb`	2560	2560
`yarn.scheduler.minimum-allocation-mb`	256	256
`yarn.scheduler.maximum-allocation-mb`	8192	15360
`yarn.nodemanager.resource.memory-mb`	30720	15360

m2.4xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx1280m	-Xmx1280m
`mapreduce.java.opts`	-Xmx2304m	-Xmx2304m
`mapreduce.map.memory.mb`	1536	1536
`mapreduce.reduce.memory.mb`	2560	2560
`yarn.app.mapreduce.am.resource.mb`	2560	2560

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	256	256
yarn.scheduler.maximum-allocation-mb	8192	30720
yarn.nodemanager.resource.memory-mb	61440	30720

m3 instances

m3.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1152m	-Xmx1152m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1440	1440
mapreduce.reduce.memory.mb	2880	2880
yarn.app.mapreduce.am.resource.mb	2880	2880
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11520	5760
yarn.nodemanager.resource.memory-mb	11520	5760

m3.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1152m	-Xmx1152m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1440	1440
mapreduce.reduce.memory.mb	2880	2880
yarn.app.mapreduce.am.resource.mb	2880	2880
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23040	11520
yarn.nodemanager.resource.memory-mb	23040	11520

m4 instances

m4.large

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072
yarn.app.mapreduce.am.resource.mb	3072	3072

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	6144	3072
yarn.nodemanager.resource.memory-mb	6144	3072

m4.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072
yarn.app.mapreduce.am.resource.mb	3072	3072
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	12288	6144
yarn.nodemanager.resource.memory-mb	12288	6144

m4.2xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx1229m	-Xmx1229m
`mapreduce.java.opts`	-Xmx2458m	-Xmx2458m
`mapreduce.map.memory.mb`	1536	1536
`mapreduce.reduce.memory.mb`	3072	3072
`yarn.app.mapreduce.am.resource.mb`	3072	3072
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	24576	12288
`yarn.nodemanager.resource.memory-mb`	24576	12288

m4.4xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx1434m	-Xmx1434m
`mapreduce.java.opts`	-Xmx2868m	-Xmx2868m
`mapreduce.map.memory.mb`	1792	1792
`mapreduce.reduce.memory.mb`	3584	3584
`yarn.app.mapreduce.am.resource.mb`	3584	3584

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

m4.10xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1557m	-Xmx1557m
mapreduce.java.opts	-Xmx3114m	-Xmx3114m
mapreduce.map.memory.mb	1946	1946
mapreduce.reduce.memory.mb	3892	3892
yarn.app.mapreduce.am.resource.mb	3892	3892
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	155648	31104
yarn.nodemanager.resource.memory-mb	155648	31104

m4.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx1587m	-Xmx1587m
mapreduce.java.opts	-Xmx3174m	-Xmx3174m
mapreduce.map.memory.mb	1984	1984
mapreduce.reduce.memory.mb	3968	3968
yarn.app.mapreduce.am.resource.mb	3968	3968
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744

m5 instances

m5.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144
yarn.app.mapreduce.am.resource.mb	6144	6144

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	12288	6144
yarn.nodemanager.resource.memory-mb	12288	6144

m5.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144
yarn.app.mapreduce.am.resource.mb	6144	6144
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

m5.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2867m	-Xmx2867m
mapreduce.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.map.memory.mb	3584	3584
mapreduce.reduce.memory.mb	7168	7168
yarn.app.mapreduce.am.resource.mb	7168	7168
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

m5.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	122880	30720
yarn.nodemanager.resource.memory-mb	122880	30720

m5.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3140m	-Xmx3140m
mapreduce.java.opts	-Xmx6280m	-Xmx6280m
mapreduce.map.memory.mb	3925	3925
mapreduce.reduce.memory.mb	7850	7850
yarn.app.mapreduce.am.resource.mb	7850	7850
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31416
yarn.nodemanager.resource.memory-mb	188416	31416

m5.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3174m	-Xmx3174m
mapreduce.java.opts	-Xmx6348m	-Xmx6348m
mapreduce.map.memory.mb	3968	3968
mapreduce.reduce.memory.mb	7936	7936
yarn.app.mapreduce.am.resource.mb	7936	7936
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744

m5.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3209m	-Xmx3209m
mapreduce.java.opts	-Xmx6418m	-Xmx6418m
mapreduce.map.memory.mb	4011	4011
mapreduce.reduce.memory.mb	8022	8022
yarn.app.mapreduce.am.resource.mb	8022	8022

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32056
yarn.nodemanager.resource.memory-mb	385024	32056

m5a instances

m5a.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144
yarn.app.mapreduce.am.resource.mb	6144	6144
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	12288	6144
yarn.nodemanager.resource.memory-mb	12288	6144

m5a.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144
yarn.app.mapreduce.am.resource.mb	6144	6144
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

m5a.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2867m	-Xmx2867m
mapreduce.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.map.memory.mb	3584	3584
mapreduce.reduce.memory.mb	7168	7168
yarn.app.mapreduce.am.resource.mb	7168	7168

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

m5a.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	122880	30720
yarn.nodemanager.resource.memory-mb	122880	30720

m5a.12xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx3140m	-Xmx3140m
`mapreduce.java.opts`	-Xmx6280m	-Xmx6280m
`mapreduce.map.memory.mb`	3925	3925
`mapreduce.reduce.memory.mb`	7850	7850
`yarn.app.mapreduce.am.resource.mb`	7850	7850
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	188416	31416
`yarn.nodemanager.resource.memory-mb`	188416	31416

m5a.16xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx3174m	-Xmx3174m
`mapreduce.java.opts`	-Xmx6348m	-Xmx6348m
`mapreduce.map.memory.mb`	3968	3968
`mapreduce.reduce.memory.mb`	7936	7936
`yarn.app.mapreduce.am.resource.mb`	7936	7936

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744

m5a.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3209m	-Xmx3209m
mapreduce.java.opts	-Xmx6418m	-Xmx6418m
mapreduce.map.memory.mb	4011	4011
mapreduce.reduce.memory.mb	8022	8022
yarn.app.mapreduce.am.resource.mb	8022	8022
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32056
yarn.nodemanager.resource.memory-mb	385024	32056

m5ad instances

m5ad.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

m5ad.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

m5ad.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

m5ad.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

m5ad.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allo cation-mb	32	32
yarn.scheduler.maximum-allo cation-mb	179200	29880
yarn.nodemanager.resource.m emory-mb	179200	29880

m5ad.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resou rce.mb	7552	7552
yarn.scheduler.minimum-allo cation-mb	32	32
yarn.scheduler.maximum-allo cation-mb	241664	30208
yarn.nodemanager.resource.m emory-mb	241664	30208

m5ad.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

m5d instances

m5d.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144
yarn.app.mapreduce.am.resource.mb	6144	6144

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	12288	6144
yarn.nodemanager.resource.memory-mb	12288	6144

m5d.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144
yarn.app.mapreduce.am.resource.mb	6144	6144
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

m5d.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2867m	-Xmx2867m
mapreduce.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.map.memory.mb	3584	3584
mapreduce.reduce.memory.mb	7168	7168
yarn.app.mapreduce.am.resource.mb	7168	7168
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

m5d.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	122880	30720
yarn.nodemanager.resource.memory-mb	122880	30720

m5d.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3140m	-Xmx3140m
mapreduce.java.opts	-Xmx6280m	-Xmx6280m
mapreduce.map.memory.mb	3925	3925
mapreduce.reduce.memory.mb	7850	7850
yarn.app.mapreduce.am.resource.mb	7850	7850
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31416
yarn.nodemanager.resource.memory-mb	188416	31416

m5d.16xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx3174m	-Xmx3174m
`mapreduce.java.opts`	-Xmx6348m	-Xmx6348m
`mapreduce.map.memory.mb`	3968	3968
`mapreduce.reduce.memory.mb`	7936	7936
`yarn.app.mapreduce.am.resource.mb`	7936	7936
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	253952	31744
`yarn.nodemanager.resource.memory-mb`	253952	31744

m5d.24xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx3209m	-Xmx3209m
`mapreduce.java.opts`	-Xmx6418m	-Xmx6418m
`mapreduce.map.memory.mb`	4011	4011
`mapreduce.reduce.memory.mb`	8022	8022
`yarn.app.mapreduce.am.resource.mb`	8022	8022

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32056
yarn.nodemanager.resource.memory-mb	385024	32056

m5dn instances

m5dn.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

m5dn.2xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx2342m	-Xmx2342m
`mapreduce.java.opts`	-Xmx4684m	-Xmx4684m
`mapreduce.map.memory.mb`	2928	2928
`mapreduce.reduce.memory.mb`	5856	5856
`yarn.app.mapreduce.am.resource.mb`	5856	5856
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	23424	11712
`yarn.nodemanager.resource.memory-mb`	23424	11712

m5dn.4xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx2714m	-Xmx2714m
`mapreduce.java.opts`	-Xmx5428m	-Xmx5428m
`mapreduce.map.memory.mb`	3392	3392
`mapreduce.reduce.memory.mb`	6784	6784
`yarn.app.mapreduce.am.resource.mb`	6784	6784

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

m5dn.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

m5dn.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

m5dn.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

m5dn.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

m5n instances

m5n.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

m5n.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

m5n.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

m5n.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

m5n.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

m5n.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

m5n.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

m5zn instances

m5zn.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.java.opts	-Xmx4608m	-Xmx4608m
mapreduce.map.memory.mb	2880	2880
mapreduce.reduce.memory.mb	5760	5760
yarn.app.mapreduce.am.resource.mb	5760	5760

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11520	5760
yarn.nodemanager.resource.memory-mb	11520	5760

m5zn.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

m5zn.3xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2577m	-Xmx2577m
mapreduce.java.opts	-Xmx5154m	-Xmx5154m
mapreduce.map.memory.mb	3221	3221
mapreduce.reduce.memory.mb	6442	6442
yarn.app.mapreduce.am.resource.mb	6442	6442
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	38656	19328
yarn.nodemanager.resource.memory-mb	38656	19328

m5zn.6xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.java.opts	-Xmx5700m	-Xmx5700m
mapreduce.map.memory.mb	3563	3563
mapreduce.reduce.memory.mb	7126	7126
yarn.app.mapreduce.am.resource.mb	7126	7126

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	28496
yarn.nodemanager.resource.memory-mb	85504	28496

m5zn.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

m6a instances

m6a.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

m6a.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

m6a.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

m6a.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

m6a.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

m6a.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

m6a.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

m6a.32xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680

Configuration option	Default value	With HBase installed
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	491520	30720
`yarn.nodemanager.resource.memory-mb`	491520	30720

m6a.48xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx3089m	-Xmx3089m
`mapreduce.java.opts`	-Xmx6178m	-Xmx6178m
`mapreduce.map.memory.mb`	3861	3861
`mapreduce.reduce.memory.mb`	7722	7722
`yarn.app.mapreduce.am.resource.mb`	7722	7722
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	741376	30952
`yarn.nodemanager.resource.memory-mb`	741376	30952

m6g instances

m6g.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

m6g.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

m6g.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

m6g.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

m6g.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	181248	30208
yarn.nodemanager.resource.memory-mb	181248	30208

m6g.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

m6gd instances

m6gd.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

m6gd.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

m6gd.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

m6gd.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

m6gd.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	181248	30208
yarn.nodemanager.resource.memory-mb	181248	30208

m6gd.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

m6i instances

m6i.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

m6i.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

m6i.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

m6i.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

m6i.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	181248	30208
yarn.nodemanager.resource.memory-mb	181248	30208

m6i.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

m6i.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

m6i.32xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

m6id instances

m6id.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

m6id.2xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx2342m	-Xmx2342m
`mapreduce.java.opts`	-Xmx4684m	-Xmx4684m
`mapreduce.map.memory.mb`	2928	2928
`mapreduce.reduce.memory.mb`	5856	5856
`yarn.app.mapreduce.am.resource.mb`	5856	5856
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	23424	11712
`yarn.nodemanager.resource.memory-mb`	23424	11712

m6id.4xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx2714m	-Xmx2714m
`mapreduce.java.opts`	-Xmx5428m	-Xmx5428m
`mapreduce.map.memory.mb`	3392	3392
`mapreduce.reduce.memory.mb`	6784	6784
`yarn.app.mapreduce.am.resource.mb`	6784	6784

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

m6id.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

m6id.12xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx2986m	-Xmx2986m
`mapreduce.java.opts`	-Xmx5972m	-Xmx5972m
`mapreduce.map.memory.mb`	3733	3733
`mapreduce.reduce.memory.mb`	7466	7466
`yarn.app.mapreduce.am.resource.mb`	7466	7466
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	179200	29880
`yarn.nodemanager.resource.memory-mb`	179200	29880

m6id.16xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx3021m	-Xmx3021m
`mapreduce.java.opts`	-Xmx6042m	-Xmx6042m
`mapreduce.map.memory.mb`	3776	3776
`mapreduce.reduce.memory.mb`	7552	7552
`yarn.app.mapreduce.am.resource.mb`	7552	7552

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

m6id.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

m6id.32xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

m6idn instances

m6idn.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

m6idn.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

m6idn.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

m6idn.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

m6idn.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

m6idn.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

m6idn.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

m6idn.32xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

m6in instances

m6in.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

m6in.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

m6in.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

m6in.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

m6in.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	181248	30208
yarn.nodemanager.resource.memory-mb	181248	30208

m6in.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

m6in.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

m6in.32xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

m7a instances

m7a.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

m7a.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

m7a.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

m7a.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

m7a.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

m7a.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

m7a.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

m7a.32xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx3072m	-Xmx3072m
`mapreduce.java.opts`	-Xmx6144m	-Xmx6144m
`mapreduce.map.memory.mb`	3840	3840
`mapreduce.reduce.memory.mb`	7680	7680
`yarn.app.mapreduce.am.resource.mb`	7680	7680
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	491520	30720
`yarn.nodemanager.resource.memory-mb`	491520	30720

m7a.48xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx3089m	-Xmx3089m
`mapreduce.java.opts`	-Xmx6178m	-Xmx6178m
`mapreduce.map.memory.mb`	3861	3861
`mapreduce.reduce.memory.mb`	7722	7722
`yarn.app.mapreduce.am.resource.mb`	7722	7722

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30952
yarn.nodemanager.resource.memory-mb	741376	30952

m7g instances

m7g.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

m7g.2xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx2342m	-Xmx2342m
`mapreduce.java.opts`	-Xmx4684m	-Xmx4684m
`mapreduce.map.memory.mb`	2928	2928
`mapreduce.reduce.memory.mb`	5856	5856
`yarn.app.mapreduce.am.resource.mb`	5856	5856
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	23424	11712
`yarn.nodemanager.resource.memory-mb`	23424	11712

m7g.4xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx2714m	-Xmx2714m
`mapreduce.java.opts`	-Xmx5428m	-Xmx5428m
`mapreduce.map.memory.mb`	3392	3392
`mapreduce.reduce.memory.mb`	6784	6784
`yarn.app.mapreduce.am.resource.mb`	6784	6784

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

m7g.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

m7g.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

m7g.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

m7gd instances

m7gd.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

m7gd.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

m7gd.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

m7gd.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

m7gd.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

m7gd.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

Configuration option	Default value	With HBase installed
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	241664	30208
`yarn.nodemanager.resource.memory-mb`	241664	30208

m7i instances

m7i.xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx2342m	-Xmx2342m
`mapreduce.java.opts`	-Xmx4684m	-Xmx4684m
`mapreduce.map.memory.mb`	2928	2928
`mapreduce.reduce.memory.mb`	5856	5856
`yarn.app.mapreduce.am.resource.mb`	5856	5856
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	11712	5856
`yarn.nodemanager.resource.memory-mb`	11712	5856

m7i.2xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx2342m	-Xmx2342m
`mapreduce.java.opts`	-Xmx4684m	-Xmx4684m
`mapreduce.map.memory.mb`	2928	2928
`mapreduce.reduce.memory.mb`	5856	5856
`yarn.app.mapreduce.am.resource.mb`	5856	5856
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	23424	11712
`yarn.nodemanager.resource.memory-mb`	23424	11712

m7i.4xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx2714m	-Xmx2714m
`mapreduce.java.opts`	-Xmx5428m	-Xmx5428m
`mapreduce.map.memory.mb`	3392	3392
`mapreduce.reduce.memory.mb`	6784	6784
`yarn.app.mapreduce.am.resource.mb`	6784	6784

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

m7i.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

m7i.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	181248	30208
yarn.nodemanager.resource.memory-mb	181248	30208

m7i.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

m7i.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

m7i.48xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3089m	-Xmx3089m
mapreduce.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.map.memory.mb	3861	3861
mapreduce.reduce.memory.mb	7722	7722
yarn.app.mapreduce.am.resource.mb	7722	7722
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30952
yarn.nodemanager.resource.memory-mb	741376	30952

m7i-flex instances

m7i-flex.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

m7i-flex.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

m7i-flex.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

m7i-flex.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

p2 instances

p2.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.java.opts	-Xmx21708m	-Xmx21708m
mapreduce.map.memory.mb	13568	13568
mapreduce.reduce.memory.mb	27136	27136
yarn.app.mapreduce.am.resource.mb	27136	27136
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

p2.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.java.opts	-Xmx24576m	-Xmx24576m
mapreduce.map.memory.mb	15360	15360
mapreduce.reduce.memory.mb	30720	30720
yarn.app.mapreduce.am.resource.mb	30720	30720
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

p2.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx9267m	-Xmx9267m
mapreduce.java.opts	-Xmx18534m	-Xmx18534m
mapreduce.map.memory.mb	11584	11584
mapreduce.reduce.memory.mb	23168	23168
yarn.app.mapreduce.am.resource.mb	23168	23168

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	23168
yarn.nodemanager.resource.memory-mb	741376	23168

p3 instances

p3.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

p3.8xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx6042m	-Xmx6042m
`mapreduce.java.opts`	-Xmx12084m	-Xmx12084m
`mapreduce.map.memory.mb`	7552	7552
`mapreduce.reduce.memory.mb`	15104	15104
`yarn.app.mapreduce.am.resource.mb`	15104	15104
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	241664	30208
`yarn.nodemanager.resource.memory-mb`	241664	30208

p3.16xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx6144m	-Xmx6144m
`mapreduce.java.opts`	-Xmx12288m	-Xmx12288m
`mapreduce.map.memory.mb`	7680	7680
`mapreduce.reduce.memory.mb`	15360	15360
`yarn.app.mapreduce.am.resource.mb`	15360	15360

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

p5 instances

p5.48xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx8294m	-Xmx8294m
mapreduce.java.opts	-Xmx16588m	-Xmx16588m
mapreduce.map.memory.mb	10368	10368
mapreduce.reduce.memory.mb	20736	20736
yarn.app.mapreduce.am.resource.mb	20736	20736
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1990656	20736
yarn.nodemanager.resource.memory-mb	1990656	20736

r3 instances

r3.xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx2342m	-Xmx2342m
`mapreduce.java.opts`	-Xmx4684m	-Xmx4684m
`mapreduce.map.memory.mb`	2928	2928
`mapreduce.reduce.memory.mb`	5856	5856
`yarn.app.mapreduce.am.resource.mb`	5856	5856
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	23424	11712
`yarn.nodemanager.resource.memory-mb`	23424	11712

r3.2xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx2714m	-Xmx2714m
`mapreduce.java.opts`	-Xmx5428m	-Xmx5428m
`mapreduce.map.memory.mb`	3392	3392
`mapreduce.reduce.memory.mb`	6784	6784
`yarn.app.mapreduce.am.resource.mb`	6784	6784

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r3.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

r3.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r4 instances

r4.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r4.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r4.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

r4.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r4.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

r5 instances

r5.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4915m	-Xmx4915m
mapreduce.java.opts	-Xmx9830m	-Xmx9830m
mapreduce.map.memory.mb	6144	6144
mapreduce.reduce.memory.mb	12288	12288
yarn.app.mapreduce.am.resource.mb	12288	12288
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

r5.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.java.opts	-Xmx11468m	-Xmx11468m
mapreduce.map.memory.mb	7168	7168
mapreduce.reduce.memory.mb	14336	14336
yarn.app.mapreduce.am.resource.mb	14336	14336

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

r5.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	122880	30720
yarn.nodemanager.resource.memory-mb	122880	30720

r5.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6349m	-Xmx6349m
mapreduce.java.opts	-Xmx12698m	-Xmx12698m
mapreduce.map.memory.mb	7936	7936
mapreduce.reduce.memory.mb	15872	15872
yarn.app.mapreduce.am.resource.mb	15872	15872
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744

r5.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6417m	-Xmx6417m
mapreduce.java.opts	-Xmx12834m	-Xmx12834m
mapreduce.map.memory.mb	8021	8021
mapreduce.reduce.memory.mb	16042	16042
yarn.app.mapreduce.am.resource.mb	16042	16042

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32100
yarn.nodemanager.resource.memory-mb	385024	32100

r5.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6451m	-Xmx6451m
mapreduce.java.opts	-Xmx12902m	-Xmx12902m
mapreduce.map.memory.mb	8064	8064
mapreduce.reduce.memory.mb	16128	16128
yarn.app.mapreduce.am.resource.mb	16128	16128
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	516096	32256
yarn.nodemanager.resource.memory-mb	516096	32256

r5.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6486m	-Xmx6486m
mapreduce.java.opts	-Xmx12972m	-Xmx12972m
mapreduce.map.memory.mb	8107	8107
mapreduce.reduce.memory.mb	16214	16214
yarn.app.mapreduce.am.resource.mb	16214	16214
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	778240	32396
yarn.nodemanager.resource.memory-mb	778240	32396

r5a instances

r5a.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4915m	-Xmx4915m
mapreduce.java.opts	-Xmx9830m	-Xmx9830m
mapreduce.map.memory.mb	6144	6144
mapreduce.reduce.memory.mb	12288	12288
yarn.app.mapreduce.am.resource.mb	12288	12288

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

r5a.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.java.opts	-Xmx11468m	-Xmx11468m
mapreduce.map.memory.mb	7168	7168
mapreduce.reduce.memory.mb	14336	14336
yarn.app.mapreduce.am.resource.mb	14336	14336
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

r5a.4xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx6144m	-Xmx6144m
`mapreduce.java.opts`	-Xmx12288m	-Xmx12288m
`mapreduce.map.memory.mb`	7680	7680
`mapreduce.reduce.memory.mb`	15360	15360
`yarn.app.mapreduce.am.resource.mb`	15360	15360
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	122880	30720
`yarn.nodemanager.resource.memory-mb`	122880	30720

r5a.8xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx6349m	-Xmx6349m
`mapreduce.java.opts`	-Xmx12698m	-Xmx12698m
`mapreduce.map.memory.mb`	7936	7936
`mapreduce.reduce.memory.mb`	15872	15872
`yarn.app.mapreduce.am.resource.mb`	15872	15872

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744

r5a.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6417m	-Xmx6417m
mapreduce.java.opts	-Xmx12834m	-Xmx12834m
mapreduce.map.memory.mb	8021	8021
mapreduce.reduce.memory.mb	16042	16042
yarn.app.mapreduce.am.resource.mb	16042	16042
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32100
yarn.nodemanager.resource.memory-mb	385024	32100

r5a.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6451m	-Xmx6451m
mapreduce.java.opts	-Xmx12902m	-Xmx12902m
mapreduce.map.memory.mb	8064	8064
mapreduce.reduce.memory.mb	16128	16128
yarn.app.mapreduce.am.resource.mb	16128	16128
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	516096	32256
yarn.nodemanager.resource.memory-mb	516096	32256

r5a.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6486m	-Xmx6486m
mapreduce.java.opts	-Xmx12972m	-Xmx12972m
mapreduce.map.memory.mb	8107	8107
mapreduce.reduce.memory.mb	16214	16214
yarn.app.mapreduce.am.resource.mb	16214	16214

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	778240	32396
yarn.nodemanager.resource.memory-mb	778240	32396

r5ad instances

r5ad.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r5ad.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r5ad.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

r5ad.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r5ad.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

r5ad.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6246m	-Xmx6246m
mapreduce.java.opts	-Xmx12492m	-Xmx12492m
mapreduce.map.memory.mb	7808	7808
mapreduce.reduce.memory.mb	15616	15616
yarn.app.mapreduce.am.resource.mb	15616	15616

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	499712	31232
yarn.nodemanager.resource.memory-mb	499712	31232

r5ad.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

r5b instances

r5b.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r5b.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r5b.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

r5b.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r5b.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

r5b.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

r5b.24xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx6178m	-Xmx6178m
`mapreduce.java.opts`	-Xmx12356m	-Xmx12356m
`mapreduce.map.memory.mb`	7723	7723
`mapreduce.reduce.memory.mb`	15446	15446
`yarn.app.mapreduce.am.resource.mb`	15446	15446
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	741376	30860
`yarn.nodemanager.resource.memory-mb`	741376	30860

r5d instances

r5d.xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx4915m	-Xmx4915m
`mapreduce.java.opts`	-Xmx9830m	-Xmx9830m
`mapreduce.map.memory.mb`	6144	6144
`mapreduce.reduce.memory.mb`	12288	12288
`yarn.app.mapreduce.am.resource.mb`	12288	12288

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

r5d.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.java.opts	-Xmx11468m	-Xmx11468m
mapreduce.map.memory.mb	7168	7168
mapreduce.reduce.memory.mb	14336	14336
yarn.app.mapreduce.am.resource.mb	14336	14336
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

r5d.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	122880	30720
yarn.nodemanager.resource.memory-mb	122880	30720

r5d.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6349m	-Xmx6349m
mapreduce.java.opts	-Xmx12698m	-Xmx12698m
mapreduce.map.memory.mb	7936	7936
mapreduce.reduce.memory.mb	15872	15872
yarn.app.mapreduce.am.resource.mb	15872	15872

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744

r5d.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6417m	-Xmx6417m
mapreduce.java.opts	-Xmx12834m	-Xmx12834m
mapreduce.map.memory.mb	8021	8021
mapreduce.reduce.memory.mb	16042	16042
yarn.app.mapreduce.am.resource.mb	16042	16042
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32100
yarn.nodemanager.resource.memory-mb	385024	32100

r5d.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6451m	-Xmx6451m
mapreduce.java.opts	-Xmx12902m	-Xmx12902m
mapreduce.map.memory.mb	8064	8064
mapreduce.reduce.memory.mb	16128	16128
yarn.app.mapreduce.am.resource.mb	16128	16128
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	516096	32256
yarn.nodemanager.resource.memory-mb	516096	32256

r5d.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6486m	-Xmx6486m
mapreduce.java.opts	-Xmx12972m	-Xmx12972m
mapreduce.map.memory.mb	8107	8107
mapreduce.reduce.memory.mb	16214	16214
yarn.app.mapreduce.am.resource.mb	16214	16214

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	778240	32396
yarn.nodemanager.resource.memory-mb	778240	32396

r5dn instances

r5dn.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r5dn.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r5dn.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

r5dn.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r5dn.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

r5dn.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

r5dn.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

r5n instances

r5n.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r5n.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r5n.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

r5n.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r5n.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

r5n.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

r5n.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

r6a instances

r6a.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r6a.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r6a.4xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx5837m	-Xmx5837m
`mapreduce.java.opts`	-Xmx11674m	-Xmx11674m
`mapreduce.map.memory.mb`	7296	7296
`mapreduce.reduce.memory.mb`	14592	14592
`yarn.app.mapreduce.am.resource.mb`	14592	14592
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	116736	29184
`yarn.nodemanager.resource.memory-mb`	116736	29184

r6a.8xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx6042m	-Xmx6042m
`mapreduce.java.opts`	-Xmx12084m	-Xmx12084m
`mapreduce.map.memory.mb`	7552	7552
`mapreduce.reduce.memory.mb`	15104	15104
`yarn.app.mapreduce.am.resource.mb`	15104	15104

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r6a.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

r6a.16xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx6144m	-Xmx6144m
`mapreduce.java.opts`	-Xmx12288m	-Xmx12288m
`mapreduce.map.memory.mb`	7680	7680
`mapreduce.reduce.memory.mb`	15360	15360
`yarn.app.mapreduce.am.resource.mb`	15360	15360
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	491520	30720
`yarn.nodemanager.resource.memory-mb`	491520	30720

r6a.24xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx6178m	-Xmx6178m
`mapreduce.java.opts`	-Xmx12356m	-Xmx12356m
`mapreduce.map.memory.mb`	7723	7723
`mapreduce.reduce.memory.mb`	15446	15446
`yarn.app.mapreduce.am.resource.mb`	15446	15446

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

r6a.32xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	7744	7744
mapreduce.reduce.memory.mb	15488	15488
yarn.app.mapreduce.am.resource.mb	15488	15488
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

r6a.48xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6212m	-Xmx6212m
mapreduce.java.opts	-Xmx12424m	-Xmx12424m
mapreduce.map.memory.mb	7765	7765
mapreduce.reduce.memory.mb	15530	15530
yarn.app.mapreduce.am.resource.mb	15530	15530
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1490944	31124
yarn.nodemanager.resource.memory-mb	1490944	31124

r6g instances

r6g.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r6g.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r6g.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

r6g.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r6g.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

r6g.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

r6gd instances

r6gd.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r6gd.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r6gd.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

r6gd.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r6gd.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

r6gd.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

r6i instances

r6i.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r6i.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r6i.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

r6i.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r6i.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

r6i.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

r6i.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

r6i.32xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6029m	-Xmx6029m
mapreduce.java.opts	-Xmx12058m	-Xmx12058m
mapreduce.map.memory.mb	7536	7536
mapreduce.reduce.memory.mb	15072	15072
yarn.app.mapreduce.am.resource.mb	15072	15072
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	964608	30144
yarn.nodemanager.resource.memory-mb	964608	30144

r6id instances

r6id.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r6id.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r6id.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

r6id.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r6id.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

r6id.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

r6id.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

r6id.32xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	7744	7744
mapreduce.reduce.memory.mb	15488	15488
yarn.app.mapreduce.am.resource.mb	15488	15488

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

r6idn instances

r6idn.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r6idn.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r6idn.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

r6idn.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r6idn.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

r6idn.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

r6idn.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

r6idn.32xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	7744	7744
mapreduce.reduce.memory.mb	15488	15488
yarn.app.mapreduce.am.resource.mb	15488	15488
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

r6in instances

r6in.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r6in.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r6in.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

r6in.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r6in.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

r6in.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

r6in.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

r6in.32xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	7744	7744
mapreduce.reduce.memory.mb	15488	15488
yarn.app.mapreduce.am.resource.mb	15488	15488
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

r7a instances

r7a.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r7a.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r7a.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

r7a.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r7a.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

r7a.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

r7a.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

r7a.32xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	7744	7744
mapreduce.reduce.memory.mb	15488	15488
yarn.app.mapreduce.am.resource.mb	15488	15488

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

r7a.48xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6212m	-Xmx6212m
mapreduce.java.opts	-Xmx12424m	-Xmx12424m
mapreduce.map.memory.mb	7765	7765
mapreduce.reduce.memory.mb	15530	15530
yarn.app.mapreduce.am.resource.mb	15530	15530
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1490944	31124
yarn.nodemanager.resource.memory-mb	1490944	31124

r7g instances

r7g.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r7g.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r7g.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

r7g.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r7g.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

r7g.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

r7gd instances

r7gd.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r7gd.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r7gd.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

r7gd.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r7gd.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

r7gd.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

r7i instances

r7i.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r7i.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r7i.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

r7i.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r7i.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

r7i.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

r7i.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

r7i.48xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6212m	-Xmx6212m
mapreduce.java.opts	-Xmx12424m	-Xmx12424m
mapreduce.map.memory.mb	7765	7765
mapreduce.reduce.memory.mb	15530	15530
yarn.app.mapreduce.am.resource.mb	15530	15530

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1490944	31124
yarn.nodemanager.resource.memory-mb	1490944	31124

r7iz instances

r7iz.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r7iz.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r7iz.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

r7iz.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r7iz.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

r7iz.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

r7iz.32xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6029m	-Xmx6029m
mapreduce.java.opts	-Xmx12058m	-Xmx12058m
mapreduce.map.memory.mb	7536	7536
mapreduce.reduce.memory.mb	15072	15072
yarn.app.mapreduce.am.resource.mb	15072	15072
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	964608	30144
yarn.nodemanager.resource.memory-mb	964608	30144

x1 instances

x1.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx12058m	-Xmx12058m
mapreduce.java.opts	-Xmx24116m	-Xmx24116m
mapreduce.map.memory.mb	15072	15072
mapreduce.reduce.memory.mb	30144	30144
yarn.app.mapreduce.am.resource.mb	30144	30144
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	964608	30144
yarn.nodemanager.resource.memory-mb	964608	30144

x1.32xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx12109m	-Xmx12109m
mapreduce.java.opts	-Xmx24218m	-Xmx24218m
mapreduce.map.memory.mb	15136	15136
mapreduce.reduce.memory.mb	30272	30272
yarn.app.mapreduce.am.resource.mb	30272	30272

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1937408	30272
yarn.nodemanager.resource.memory-mb	1937408	30272

x1e instances

x1e.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx22682m	-Xmx22682m
mapreduce.java.opts	-Xmx45364m	-Xmx45364m
mapreduce.map.memory.mb	28352	28352
mapreduce.reduce.memory.mb	56704	56704
yarn.app.mapreduce.am.resource.mb	56704	56704
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	113408	0
yarn.nodemanager.resource.memory-mb	113408	0

x1e.2xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx23501m	-Xmx23501m
`mapreduce.java.opts`	-Xmx47002m	-Xmx47002m
`mapreduce.map.memory.mb`	29376	29376
`mapreduce.reduce.memory.mb`	58752	58752
`yarn.app.mapreduce.am.resource.mb`	58752	58752
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	235008	0
`yarn.nodemanager.resource.memory-mb`	235008	0

x1e.4xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx23910m	-Xmx23910m
`mapreduce.java.opts`	-Xmx47820m	-Xmx47820m
`mapreduce.map.memory.mb`	29888	29888
`mapreduce.reduce.memory.mb`	59776	59776
`yarn.app.mapreduce.am.resource.mb`	59776	59776

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	478208	0
yarn.nodemanager.resource.memory-mb	478208	0

x1e.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx24115m	-Xmx24115m
mapreduce.java.opts	-Xmx48230m	-Xmx48230m
mapreduce.map.memory.mb	30144	30144
mapreduce.reduce.memory.mb	60288	60288
yarn.app.mapreduce.am.resource.mb	60288	60288
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	964608	0
yarn.nodemanager.resource.memory-mb	964608	0

x1e.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx24218m	-Xmx24218m
mapreduce.java.opts	-Xmx48436m	-Xmx48436m
mapreduce.map.memory.mb	30272	30272
mapreduce.reduce.memory.mb	60544	60544
yarn.app.mapreduce.am.resource.mb	60544	60544
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1937408	0
yarn.nodemanager.resource.memory-mb	1937408	0

x1e.32xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx24269m	-Xmx24269m
mapreduce.java.opts	-Xmx48538m	-Xmx48538m
mapreduce.map.memory.mb	30336	30336
mapreduce.reduce.memory.mb	60672	60672
yarn.app.mapreduce.am.resource.mb	60672	60672

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	3883008	0
yarn.nodemanager.resource.memory-mb	3883008	0

x2gd instances

x2gd.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.java.opts	-Xmx21708m	-Xmx21708m
mapreduce.map.memory.mb	13568	13568
mapreduce.reduce.memory.mb	27136	27136
yarn.app.mapreduce.am.resource.mb	27136	27136
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

x2gd.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.java.opts	-Xmx23348m	-Xmx23348m
mapreduce.map.memory.mb	14592	14592
mapreduce.reduce.memory.mb	29184	29184
yarn.app.mapreduce.am.resource.mb	29184	29184
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

x2gd.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx12083m	-Xmx12083m
mapreduce.java.opts	-Xmx24166m	-Xmx24166m
mapreduce.map.memory.mb	15104	15104
mapreduce.reduce.memory.mb	30208	30208
yarn.app.mapreduce.am.resource.mb	30208	30208

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

x2gd.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.java.opts	-Xmx24576m	-Xmx24576m
mapreduce.map.memory.mb	15360	15360
mapreduce.reduce.memory.mb	30720	30720
yarn.app.mapreduce.am.resource.mb	30720	30720
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

x2gd.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.java.opts	-Xmx24712m	-Xmx24712m
mapreduce.map.memory.mb	15445	15445
mapreduce.reduce.memory.mb	30890	30890
yarn.app.mapreduce.am.resource.mb	30890	30890
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30906
yarn.nodemanager.resource.memory-mb	741376	30906

x2gd.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.java.opts	-Xmx24780m	-Xmx24780m
mapreduce.map.memory.mb	15488	15488
mapreduce.reduce.memory.mb	30976	30976
yarn.app.mapreduce.am.resource.mb	30976	30976

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

x2idn instances

x2idn.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.java.opts	-Xmx24780m	-Xmx24780m
mapreduce.map.memory.mb	15488	15488
mapreduce.reduce.memory.mb	30976	30976
yarn.app.mapreduce.am.resource.mb	30976	30976
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

x2idn.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx12425m	-Xmx12425m
mapreduce.java.opts	-Xmx24850m	-Xmx24850m
mapreduce.map.memory.mb	15531	15531
mapreduce.reduce.memory.mb	31062	31062
yarn.app.mapreduce.am.resource.mb	31062	31062
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1490944	31030
yarn.nodemanager.resource.memory-mb	1490944	31030

x2idn.32xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx12442m	-Xmx12442m
mapreduce.java.opts	-Xmx24884m	-Xmx24884m
mapreduce.map.memory.mb	15552	15552
mapreduce.reduce.memory.mb	31104	31104
yarn.app.mapreduce.am.resource.mb	31104	31104

Configuration option	Default value	With HBase installed
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	1990656	31104
`yarn.nodemanager.resource.memory-mb`	1990656	31104

x2iedn instances

x2iedn.xlarge

Configuration option	Default value	With HBase installed
`mapreduce.map.java.opts`	-Xmx23347m	-Xmx23347m
`mapreduce.java.opts`	-Xmx46694m	-Xmx46694m
`mapreduce.map.memory.mb`	29184	29184
`mapreduce.reduce.memory.mb`	58368	58368
`yarn.app.mapreduce.am.resource.mb`	58368	58368
`yarn.scheduler.minimum-allocation-mb`	32	32
`yarn.scheduler.maximum-allocation-mb`	116736	0
`yarn.nodemanager.resource.memory-mb`	116736	0

x2iedn.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx24166m	-Xmx24166m
mapreduce.java.opts	-Xmx48332m	-Xmx48332m
mapreduce.map.memory.mb	30208	30208
mapreduce.reduce.memory.mb	60416	60416
yarn.app.mapreduce.am.resource.mb	60416	60416
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	0
yarn.nodemanager.resource.memory-mb	241664	0

x2iedn.4xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx24576m	-Xmx24576m
mapreduce.java.opts	-Xmx49152m	-Xmx49152m
mapreduce.map.memory.mb	30720	30720
mapreduce.reduce.memory.mb	61440	61440
yarn.app.mapreduce.am.resource.mb	61440	61440

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	0
yarn.nodemanager.resource.memory-mb	491520	0

x2iedn.8xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx24781m	-Xmx24781m
mapreduce.java.opts	-Xmx49562m	-Xmx49562m
mapreduce.map.memory.mb	30976	30976
mapreduce.reduce.memory.mb	61952	61952
yarn.app.mapreduce.am.resource.mb	61952	61952
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	0
yarn.nodemanager.resource.memory-mb	991232	0

x2iedn.16xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx24883m	-Xmx24883m
mapreduce.java.opts	-Xmx49766m	-Xmx49766m
mapreduce.map.memory.mb	31104	31104
mapreduce.reduce.memory.mb	62208	62208
yarn.app.mapreduce.am.resource.mb	62208	62208
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1990656	0
yarn.nodemanager.resource.memory-mb	1990656	0

x2iedn.24xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx24918m	-Xmx24918m
mapreduce.java.opts	-Xmx49836m	-Xmx49836m
mapreduce.map.memory.mb	31147	31147
mapreduce.reduce.memory.mb	62294	62294
yarn.app.mapreduce.am.resource.mb	62294	62294

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	2990080	-32
yarn.nodemanager.resource.memory-mb	2990080	-32

x2iedn.32xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx24934m	-Xmx24934m
mapreduce.java.opts	-Xmx49868m	-Xmx49868m
mapreduce.map.memory.mb	31168	31168
mapreduce.reduce.memory.mb	62336	62336
yarn.app.mapreduce.am.resource.mb	62336	62336
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	3989504	0
yarn.nodemanager.resource.memory-mb	3989504	0

z1d instances

z1d.xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx4915m	-Xmx4915m
mapreduce.java.opts	-Xmx9830m	-Xmx9830m
mapreduce.map.memory.mb	6144	6144
mapreduce.reduce.memory.mb	12288	12288
yarn.app.mapreduce.am.resource.mb	12288	12288
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

z1d.2xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.java.opts	-Xmx11468m	-Xmx11468m
mapreduce.map.memory.mb	7168	7168
mapreduce.reduce.memory.mb	14336	14336
yarn.app.mapreduce.am.resource.mb	14336	14336

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

z1d.3xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6007m	-Xmx6007m
mapreduce.java.opts	-Xmx12014m	-Xmx12014m
mapreduce.map.memory.mb	7509	7509
mapreduce.reduce.memory.mb	15018	15018
yarn.app.mapreduce.am.resource.mb	15018	15018
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	90112	30040
yarn.nodemanager.resource.memory-mb	90112	30040

z1d.6xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6281m	-Xmx6281m
mapreduce.java.opts	-Xmx12562m	-Xmx12562m
mapreduce.map.memory.mb	7851	7851
mapreduce.reduce.memory.mb	15702	15702
yarn.app.mapreduce.am.resource.mb	15702	15702
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31396
yarn.nodemanager.resource.memory-mb	188416	31396

z1d.12xlarge

Configuration option	Default value	With HBase installed
mapreduce.map.java.opts	-Xmx6417m	-Xmx6417m
mapreduce.java.opts	-Xmx12834m	-Xmx12834m
mapreduce.map.memory.mb	8021	8021
mapreduce.reduce.memory.mb	16042	16042
yarn.app.mapreduce.am.resource.mb	16042	16042

Configuration option	Default value	With HBase installed
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32100
yarn.nodemanager.resource.memory-mb	385024	32100

Hadoop daemon configuration settings

Hadoop daemon settings are different depending on the EC2 instance type that a cluster node uses. The following tables list the default configuration settings for each EC2 instance type.

To customize these settings, use the `hadoop-env` configuration classification. For more information, see Configure applications.

Instance Types

- c1 instances
- c3 instances
- c4 instances
- c5 instances
- c5a instances
- c5ad instances
- c5d instances
- c5n instances
- c6a instances
- c6g instances
- c6gd instances
- c6gn instances
- c6i instances
- c6id instances

- c6in instances
- c7a instances
- c7g instances
- c7gd instances
- c7gn instances
- c7i instances
- d2 instances
- d3 instances
- d3en instances
- g3 instances
- g3s instances
- g4dn instances
- g5 instances
- h1 instances
- i2 instances
- i3 instances
- i3en instances
- i4g instances
- i4i instances
- im4gn instances
- is4gen instances
- m1 instances
- m2 instances
- m3 instances
- m4 instances
- m5 instances
- m5a instances
- m5ad instances
- m5d instances

- m5dn instances
- m5n instances
- m5zn instances
- m6a instances
- m6g instances
- m6gd instances
- m6i instances
- m6id instances
- m6idn instances
- m6in instances
- m7a instances
- m7g instances
- m7gd instances
- m7i instances
- m7i-flex instances
- p2 instances
- p3 instances
- p5 instances
- r3 instances
- r4 instances
- r5 instances
- r5a instances
- r5ad instances
- r5b instances
- r5d instances
- r5dn instances
- r5n instances
- r6a instances
- r6g instances

c1 instances

c1.medium

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	192
YARN_PROXYSERVER_HEAPSIZE	96
YARN_NODEMANAGER_HEAPSIZE	128
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	128
HADOOP_NAMENODE_HEAPSIZE	192

Parameter	Value
HADOOP_DATANODE_HEAPSIZE	96

c1.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	768
YARN_PROXYSERVER_HEAPSIZE	384
YARN_NODEMANAGER_HEAPSIZE	512
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	512
HADOOP_NAMENODE_HEAPSIZE	768
HADOOP_DATANODE_HEAPSIZE	384

c3 instances

c3.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972

Parameter	Value
HADOOP_DATANODE_HEAPSIZE	588

c3.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2396
YARN_PROXYSERVER_HEAPSIZE	2396
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2396
HADOOP_NAMENODE_HEAPSIZE	1740
HADOOP_DATANODE_HEAPSIZE	757

c3.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2703
YARN_PROXYSERVER_HEAPSIZE	2703
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2703
HADOOP_NAMENODE_HEAPSIZE	3276
HADOOP_DATANODE_HEAPSIZE	1064

c3.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3317
YARN_PROXYSERVER_HEAPSIZE	3317
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3317
HADOOP_NAMENODE_HEAPSIZE	6348
HADOOP_DATANODE_HEAPSIZE	1679

c4 instances

c4.large

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	1152
YARN_PROXYSERVER_HEAPSIZE	1152
YARN_NODEMANAGER_HEAPSIZE	1152
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	1152
HADOOP_NAMENODE_HEAPSIZE	576
HADOOP_DATANODE_HEAPSIZE	384

c4.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c4.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2396
YARN_PROXYSERVER_HEAPSIZE	2396
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2396
HADOOP_NAMENODE_HEAPSIZE	1740
HADOOP_DATANODE_HEAPSIZE	757

c4.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2703
YARN_PROXYSERVER_HEAPSIZE	2703
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2703
HADOOP_NAMENODE_HEAPSIZE	3276
HADOOP_DATANODE_HEAPSIZE	1064

c4.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3317
YARN_PROXYSERVER_HEAPSIZE	3317
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3317
HADOOP_NAMENODE_HEAPSIZE	6348
HADOOP_DATANODE_HEAPSIZE	1679

c5 instances

c5.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2252
YARN_PROXYSERVER_HEAPSIZE	2252
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2252
HADOOP_NAMENODE_HEAPSIZE	1024
HADOOP_DATANODE_HEAPSIZE	614

c5.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

c5.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

c5.9xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3563
YARN_PROXYSERVER_HEAPSIZE	3563
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3563
HADOOP_NAMENODE_HEAPSIZE	7577
HADOOP_DATANODE_HEAPSIZE	1925

c5.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

c5.18xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5038
YARN_PROXYSERVER_HEAPSIZE	5038
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5038
HADOOP_NAMENODE_HEAPSIZE	14950
HADOOP_DATANODE_HEAPSIZE	3399

c5.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

c5a instances

c5a.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c5a.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c5a.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c5a.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c5a.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

c5a.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

c5a.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

c5ad instances

c5ad.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c5ad.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c5ad.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c5ad.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c5ad.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c5ad.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

c5ad.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

c5d instances

c5d.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2252
YARN_PROXYSERVER_HEAPSIZE	2252
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2252
HADOOP_NAMENODE_HEAPSIZE	1024
HADOOP_DATANODE_HEAPSIZE	614

c5d.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

c5d.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

c5d.9xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3563
YARN_PROXYSERVER_HEAPSIZE	3563
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3563
HADOOP_NAMENODE_HEAPSIZE	7577
HADOOP_DATANODE_HEAPSIZE	1925

c5d.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

c5d.18xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5038
YARN_PROXYSERVER_HEAPSIZE	5038
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5038
HADOOP_NAMENODE_HEAPSIZE	14950
HADOOP_DATANODE_HEAPSIZE	3399

c5d.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

c5n instances

c5n.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2304
YARN_PROXYSERVER_HEAPSIZE	2304
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2304
HADOOP_NAMENODE_HEAPSIZE	1280
HADOOP_DATANODE_HEAPSIZE	665

c5n.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2519
YARN_PROXYSERVER_HEAPSIZE	2519
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2519
HADOOP_NAMENODE_HEAPSIZE	2355
HADOOP_DATANODE_HEAPSIZE	880

c5n.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2949
YARN_PROXYSERVER_HEAPSIZE	2949
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2949
HADOOP_NAMENODE_HEAPSIZE	4505
HADOOP_DATANODE_HEAPSIZE	1310

c5n.9xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

c5n.18xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

c6a instances

c6a.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c6a.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c6a.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c6a.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c6a.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c6a.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

c6a.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

c6a.32xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

c6a.48xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

c6g instances

c6g.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c6g.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c6g.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c6g.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c6g.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c6g.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

c6gd instances

c6gd.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c6gd.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c6gd.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c6gd.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c6gd.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c6gd.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

c6gn instances

c6gn.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c6gn.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c6gn.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c6gn.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c6gn.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c6gn.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

c6i instances

c6i.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c6i.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c6i.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c6i.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c6i.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c6i.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

c6i.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

c6i.32xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

c6id instances

c6id.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c6id.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c6id.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c6id.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c6id.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c6id.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

c6id.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

c6id.32xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

c6in instances

c6in.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c6in.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c6in.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c6in.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c6in.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c6in.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

c6in.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

c6in.32xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

c7a instances

c7a.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c7a.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c7a.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c7a.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c7a.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c7a.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

c7a.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

c7a.32xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

c7a.48xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

c7g instances

c7g.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c7g.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c7g.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c7g.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c7g.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c7g.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

c7gd instances

c7gd.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c7gd.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c7gd.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c7gd.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c7gd.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c7gd.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

c7gn instances

c7gn.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c7gn.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c7gn.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c7gn.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c7gn.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c7gn.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

c7i instances

c7i.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c7i.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c7i.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c7i.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c7i.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c7i.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

c7i.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

c7i.48xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

d2 instances

d2.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

d2.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

d2.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

d2.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

d3 instances

d3.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

d3.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

d3.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

d3.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

d3en instances

d3en.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

d3en.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

d3en.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

d3en.6xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

d3en.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

d3en.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

g3 instances

g3.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

g3.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

g3.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

g3s instances

g3s.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

g4dn instances

g4dn.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

g4dn.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

g4dn.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

g4dn.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

g4dn.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

g4dn.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

g5 instances

g5.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

g5.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

g5.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

g5.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

g5.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

g5.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

g5.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

g5.48xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

h1 instances

h1.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

h1.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

h1.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

h1.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

i2 instances

i2.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

i2.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

i2.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

i2.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

i3 instances

i3.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

i3.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

i3.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

i3.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

i3.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

i3en instances

i3en.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

i3en.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

i3en.3xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

i3en.6xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

i3en.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

i3en.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	17817
YARN_PROXYSERVER_HEAPSIZE	17817
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17817
HADOOP_NAMENODE_HEAPSIZE	78848
HADOOP_DATANODE_HEAPSIZE	4096

i4g instances

i4g.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

i4g.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

i4g.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

i4g.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

i4g.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

i4i instances

i4i.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

i4i.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

i4i.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

i4i.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

i4i.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

i4i.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

i4i.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

i4i.32xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

im4gn instances

im4gn.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

im4gn.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

im4gn.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

im4gn.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

im4gn.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

is4gen instances

is4gen.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2557
YARN_PROXYSERVER_HEAPSIZE	2557
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2557
HADOOP_NAMENODE_HEAPSIZE	2547
HADOOP_DATANODE_HEAPSIZE	919

is4gen.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3025
YARN_PROXYSERVER_HEAPSIZE	3025
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3025
HADOOP_NAMENODE_HEAPSIZE	4889
HADOOP_DATANODE_HEAPSIZE	1387

is4gen.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

is4gen.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

m1 instances

m1.small

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	256
YARN_PROXYSERVER_HEAPSIZE	96
YARN_NODEMANAGER_HEAPSIZE	192
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	128
HADOOP_NAMENODE_HEAPSIZE	192
HADOOP_DATANODE_HEAPSIZE	96

m1.medium

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	384
YARN_PROXYSERVER_HEAPSIZE	192
YARN_NODEMANAGER_HEAPSIZE	256
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	256
HADOOP_NAMENODE_HEAPSIZE	384
HADOOP_DATANODE_HEAPSIZE	192

m1.large

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	768
YARN_PROXYSERVER_HEAPSIZE	384
YARN_NODEMANAGER_HEAPSIZE	512
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	512
HADOOP_NAMENODE_HEAPSIZE	768
HADOOP_DATANODE_HEAPSIZE	384

m1.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	1024
YARN_PROXYSERVER_HEAPSIZE	512
YARN_NODEMANAGER_HEAPSIZE	768
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	1024
HADOOP_NAMENODE_HEAPSIZE	2304
HADOOP_DATANODE_HEAPSIZE	384

m2 instances

m2.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	1536
YARN_PROXYSERVER_HEAPSIZE	1024
YARN_NODEMANAGER_HEAPSIZE	1024
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	1024
HADOOP_NAMENODE_HEAPSIZE	3072
HADOOP_DATANODE_HEAPSIZE	384

m2.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	1536
YARN_PROXYSERVER_HEAPSIZE	1024
YARN_NODEMANAGER_HEAPSIZE	1024
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	1536
HADOOP_NAMENODE_HEAPSIZE	6144
HADOOP_DATANODE_HEAPSIZE	384

m2.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2048
YARN_PROXYSERVER_HEAPSIZE	1024
YARN_NODEMANAGER_HEAPSIZE	1536
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	1536
HADOOP_NAMENODE_HEAPSIZE	12288
HADOOP_DATANODE_HEAPSIZE	384

m3 instances

m3.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2396
YARN_PROXYSERVER_HEAPSIZE	2396
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2396
HADOOP_NAMENODE_HEAPSIZE	1740
HADOOP_DATANODE_HEAPSIZE	757

m3.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2703
YARN_PROXYSERVER_HEAPSIZE	2703
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2703
HADOOP_NAMENODE_HEAPSIZE	3276
HADOOP_DATANODE_HEAPSIZE	1064

m4 instances

m4.large

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2252
YARN_PROXYSERVER_HEAPSIZE	2252
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2252
HADOOP_NAMENODE_HEAPSIZE	1024
HADOOP_DATANODE_HEAPSIZE	614

m4.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

m4.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

m4.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

m4.10xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5365
YARN_PROXYSERVER_HEAPSIZE	5365
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5365
HADOOP_NAMENODE_HEAPSIZE	16588
HADOOP_DATANODE_HEAPSIZE	3727

m4.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

m5 instances

m5.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

m5.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

m5.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

m5.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

m5.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

m5.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

m5.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

m5a instances

m5a.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

m5a.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

m5a.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

m5a.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

m5a.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

m5a.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

m5a.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

m5ad instances

m5ad.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m5ad.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m5ad.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m5ad.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m5ad.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

m5ad.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m5ad.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

m5d instances

m5d.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

m5d.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

m5d.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

m5d.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

m5d.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

m5d.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

m5d.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

m5dn instances

m5dn.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m5dn.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m5dn.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m5dn.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m5dn.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

m5dn.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m5dn.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

m5n instances

m5n.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m5n.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m5n.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m5n.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m5n.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

m5n.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m5n.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

m5zn instances

m5zn.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2396
YARN_PROXYSERVER_HEAPSIZE	2396
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2396
HADOOP_NAMENODE_HEAPSIZE	1740
HADOOP_DATANODE_HEAPSIZE	757

m5zn.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m5zn.3xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3025
YARN_PROXYSERVER_HEAPSIZE	3025
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3025
HADOOP_NAMENODE_HEAPSIZE	4889
HADOOP_DATANODE_HEAPSIZE	1387

m5zn.6xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

m5zn.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

m6a instances

m6a.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m6a.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m6a.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m6a.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m6a.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

m6a.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m6a.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

m6a.32xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

m6a.48xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

m6g instances

m6g.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m6g.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m6g.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m6g.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m6g.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5877
YARN_PROXYSERVER_HEAPSIZE	5877
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5877
HADOOP_NAMENODE_HEAPSIZE	19148
HADOOP_DATANODE_HEAPSIZE	4096

m6g.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m6gd instances

m6gd.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m6gd.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m6gd.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m6gd.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m6gd.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5877
YARN_PROXYSERVER_HEAPSIZE	5877
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5877
HADOOP_NAMENODE_HEAPSIZE	19148
HADOOP_DATANODE_HEAPSIZE	4096

m6gd.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m6i instances

m6i.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m6i.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m6i.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m6i.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m6i.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5877
YARN_PROXYSERVER_HEAPSIZE	5877
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5877
HADOOP_NAMENODE_HEAPSIZE	19148
HADOOP_DATANODE_HEAPSIZE	4096

m6i.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m6i.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

m6i.32xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

m6id instances

m6id.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m6id.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m6id.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m6id.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m6id.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

m6id.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m6id.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

m6id.32xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

m6idn instances

m6idn.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m6idn.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m6idn.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m6idn.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m6idn.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

m6idn.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m6idn.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

m6idn.32xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

m6in instances

m6in.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m6in.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m6in.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m6in.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m6in.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5877
YARN_PROXYSERVER_HEAPSIZE	5877
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5877
HADOOP_NAMENODE_HEAPSIZE	19148
HADOOP_DATANODE_HEAPSIZE	4096

m6in.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m6in.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

m6in.32xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

m7a instances

m7a.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m7a.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m7a.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m7a.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m7a.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

m7a.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m7a.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

m7a.32xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

m7a.48xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

m7g instances

m7g.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m7g.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m7g.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m7g.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m7g.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

m7g.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m7gd instances

m7gd.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m7gd.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m7gd.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m7gd.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m7gd.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

m7gd.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m7i instances

m7i.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m7i.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m7i.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m7i.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m7i.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	5877
YARN_PROXYSERVER_HEAPSIZE	5877
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5877
HADOOP_NAMENODE_HEAPSIZE	19148
HADOOP_DATANODE_HEAPSIZE	4096

m7i.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m7i.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

m7i.48xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

m7i-flex instances

m7i-flex.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m7i-flex.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m7i-flex.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m7i-flex.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

p2 instances

p2.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

p2.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

p2.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

p3 instances

p3.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

p3.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

p3.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

p5 instances

p5.48xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	42065
YARN_PROXYSERVER_HEAPSIZE	42065
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	42065
HADOOP_NAMENODE_HEAPSIZE	200089
HADOOP_DATANODE_HEAPSIZE	4096

r3 instances

r3.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r3.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r3.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r3.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r4 instances

r4.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r4.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r4.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r4.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r4.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r5 instances

r5.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

r5.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

r5.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

r5.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

r5.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

r5.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12574
YARN_PROXYSERVER_HEAPSIZE	12574
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12574
HADOOP_NAMENODE_HEAPSIZE	52633
HADOOP_DATANODE_HEAPSIZE	4096

r5.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	17817
YARN_PROXYSERVER_HEAPSIZE	17817
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17817
HADOOP_NAMENODE_HEAPSIZE	78848
HADOOP_DATANODE_HEAPSIZE	4096

r5a instances

r5a.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

r5a.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

r5a.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

r5a.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

r5a.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

r5a.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12574
YARN_PROXYSERVER_HEAPSIZE	12574
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12574
HADOOP_NAMENODE_HEAPSIZE	52633
HADOOP_DATANODE_HEAPSIZE	4096

r5a.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	17817
YARN_PROXYSERVER_HEAPSIZE	17817
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17817
HADOOP_NAMENODE_HEAPSIZE	78848
HADOOP_DATANODE_HEAPSIZE	4096

r5ad instances

r5ad.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r5ad.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r5ad.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r5ad.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r5ad.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r5ad.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12247
YARN_PROXYSERVER_HEAPSIZE	12247
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12247
HADOOP_NAMENODE_HEAPSIZE	50995
HADOOP_DATANODE_HEAPSIZE	4096

r5ad.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

r5b instances

r5b.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r5b.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r5b.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r5b.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r5b.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r5b.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r5b.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

r5d instances

r5d.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

r5d.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

r5d.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

r5d.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

r5d.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

r5d.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12574
YARN_PROXYSERVER_HEAPSIZE	12574
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12574
HADOOP_NAMENODE_HEAPSIZE	52633
HADOOP_DATANODE_HEAPSIZE	4096

r5d.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	17817
YARN_PROXYSERVER_HEAPSIZE	17817
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17817
HADOOP_NAMENODE_HEAPSIZE	78848
HADOOP_DATANODE_HEAPSIZE	4096

r5dn instances

r5dn.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r5dn.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r5dn.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r5dn.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r5dn.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r5dn.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r5dn.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

r5n instances

r5n.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r5n.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r5n.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r5n.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r5n.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r5n.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r5n.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

r6a instances

r6a.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r6a.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r6a.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r6a.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r6a.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r6a.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r6a.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

r6a.32xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

r6a.48xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	32071
YARN_PROXYSERVER_HEAPSIZE	32071
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	32071
HADOOP_NAMENODE_HEAPSIZE	150118
HADOOP_DATANODE_HEAPSIZE	4096

r6g instances

r6g.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r6g.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r6g.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r6g.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r6g.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r6g.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r6gd instances

r6gd.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r6gd.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r6gd.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r6gd.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r6gd.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r6gd.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r6i instances

r6i.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r6i.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r6i.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r6i.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r6i.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r6i.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r6i.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

r6i.32xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	21544
YARN_PROXYSERVER_HEAPSIZE	21544
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	21544
HADOOP_NAMENODE_HEAPSIZE	97484
HADOOP_DATANODE_HEAPSIZE	4096

r6id instances

r6id.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r6id.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r6id.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r6id.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r6id.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r6id.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r6id.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

r6id.32xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

r6idn instances

r6idn.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r6idn.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r6idn.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r6idn.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r6idn.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r6idn.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r6idn.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

r6idn.32xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

r6in instances

r6in.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r6in.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r6in.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r6in.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r6in.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r6in.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r6in.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

r6in.32xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

r7a instances

r7a.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r7a.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r7a.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r7a.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r7a.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r7a.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r7a.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

r7a.32xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

r7a.48xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	32071
YARN_PROXYSERVER_HEAPSIZE	32071
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	32071
HADOOP_NAMENODE_HEAPSIZE	150118
HADOOP_DATANODE_HEAPSIZE	4096

r7g instances

r7g.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r7g.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r7g.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r7g.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r7g.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r7g.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r7gd instances

r7gd.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r7gd.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r7gd.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r7gd.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r7gd.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r7gd.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r7i instances

r7i.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r7i.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r7i.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r7i.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r7i.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r7i.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r7i.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

r7i.48xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	32071
YARN_PROXYSERVER_HEAPSIZE	32071
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	32071
HADOOP_NAMENODE_HEAPSIZE	150118
HADOOP_DATANODE_HEAPSIZE	4096

r7iz instances

r7iz.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r7iz.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r7iz.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r7iz.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r7iz.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r7iz.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r7iz.32xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	21544
YARN_PROXYSERVER_HEAPSIZE	21544
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	21544
HADOOP_NAMENODE_HEAPSIZE	97484
HADOOP_DATANODE_HEAPSIZE	4096

x1 instances

x1.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	21544
YARN_PROXYSERVER_HEAPSIZE	21544
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	21544
HADOOP_NAMENODE_HEAPSIZE	97484
HADOOP_DATANODE_HEAPSIZE	4096

x1.32xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	41000
YARN_PROXYSERVER_HEAPSIZE	41000
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	41000
HADOOP_NAMENODE_HEAPSIZE	194764
HADOOP_DATANODE_HEAPSIZE	4096

x1e instances

x1e.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4520
YARN_PROXYSERVER_HEAPSIZE	4520
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4520
HADOOP_NAMENODE_HEAPSIZE	12364
HADOOP_DATANODE_HEAPSIZE	2882

x1e.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	6952
YARN_PROXYSERVER_HEAPSIZE	6952
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6952
HADOOP_NAMENODE_HEAPSIZE	24524
HADOOP_DATANODE_HEAPSIZE	4096

x1e.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	11816
YARN_PROXYSERVER_HEAPSIZE	11816
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	11816
HADOOP_NAMENODE_HEAPSIZE	48844
HADOOP_DATANODE_HEAPSIZE	4096

x1e.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	21544
YARN_PROXYSERVER_HEAPSIZE	21544
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	21544
HADOOP_NAMENODE_HEAPSIZE	97484
HADOOP_DATANODE_HEAPSIZE	4096

x1e.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	41000
YARN_PROXYSERVER_HEAPSIZE	41000
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	41000
HADOOP_NAMENODE_HEAPSIZE	194764
HADOOP_DATANODE_HEAPSIZE	4096

x1e.32xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	79912
YARN_PROXYSERVER_HEAPSIZE	79912
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	79912
HADOOP_NAMENODE_HEAPSIZE	389324
HADOOP_DATANODE_HEAPSIZE	4096

x2gd instances

x2gd.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

x2gd.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

x2gd.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

x2gd.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

x2gd.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

x2gd.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

x2idn instances

x2idn.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

x2idn.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	32071
YARN_PROXYSERVER_HEAPSIZE	32071
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	32071
HADOOP_NAMENODE_HEAPSIZE	150118
HADOOP_DATANODE_HEAPSIZE	4096

x2idn.32xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	42065
YARN_PROXYSERVER_HEAPSIZE	42065
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	42065
HADOOP_NAMENODE_HEAPSIZE	200089
HADOOP_DATANODE_HEAPSIZE	4096

x2iedn instances

x2iedn.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

x2iedn.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

x2iedn.4xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

x2iedn.8xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

x2iedn.16xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	42065
YARN_PROXYSERVER_HEAPSIZE	42065
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	42065
HADOOP_NAMENODE_HEAPSIZE	200089
HADOOP_DATANODE_HEAPSIZE	4096

x2iedn.24xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	62054
YARN_PROXYSERVER_HEAPSIZE	62054
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	62054
HADOOP_NAMENODE_HEAPSIZE	300032
HADOOP_DATANODE_HEAPSIZE	4096

x2iedn.32xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	82042
YARN_PROXYSERVER_HEAPSIZE	82042
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	82042
HADOOP_NAMENODE_HEAPSIZE	399974
HADOOP_DATANODE_HEAPSIZE	4096

z1d instances

z1d.xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

z1d.2xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

z1d.3xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

z1d.6xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

z1d.12xlarge

Parameter	Value
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

HDFS configuration

The following table describes the default Hadoop Distributed File System (HDFS) parameters and their settings. You can change these values using the `hdfs-site` configuration classification. For more information, see Configure applications.

> **Warning**
>
> 1. Setting `dfs.replication` to 1 on clusters with fewer than four nodes can lead to HDFS data loss if a single node goes down. If your cluster has HDFS storage, we recommend that you configure the cluster with at least four core nodes for production workloads to avoid data loss.
> 2. Amazon EMR will not allow clusters to scale core nodes below `dfs.replication`. For example, if `dfs.replication = 2`, the minimum number of core nodes is 2.
> 3. When you use managed Scaling, auto-scaling, or choose to manually resize your cluster, we recommend that you set `dfs.replication` to 2 or higher.

Parameter	Definition	Default value
dfs.block .size	The size of HDFS blocks. When operating on data stored in HDFS, the split size is generally the size of an HDFS block. Larger numbers provide less task granularity, but also put less strain on the cluster NameNode.	134217728 (128 MB)
dfs.repli cation	The number of copies of each block to store for durability. Amazon EMR sets this value based on the number of core nodes the cluster is provisioned with. Adjust the value to meet your needs. To overwrite the default value, use the hdfs-site classification.	1 for clusters that are provisioned with less than four core nodes 2 for clusters that are provisioned with less than ten core nodes 3 for all other clusters

Transparent encryption in HDFS on Amazon EMR

Transparent encryption is implemented through the use of HDFS *encryption zones*, which are HDFS paths that you define. Each encryption zone has its own key, which is stored in the key server specified using the `hdfs-site` configuration classification.

Beginning with Amazon EMR release version 4.8.0, you can use Amazon EMR security configurations to configure data encryption settings for clusters more easily. Security configurations offer settings to enable security for data in-transit and data at-rest in Amazon Elastic Block Store (Amazon EBS) storage volumes and EMRFS data in Amazon S3. For more information, see Encrypt data in transit and at rest in the *Amazon EMR Management Guide*.

Amazon EMR uses the Hadoop KMS by default; however, you can use another KMS that implements the KeyProvider API operation. Each file in an HDFS encryption zone has its own unique *data encryption key*, which is encrypted by the encryption zone key. HDFS data is encrypted end-to-end (at-rest and in-transit) when data is written to an encryption zone because encryption and decryption activities only occur in the client.

You cannot move files between encryptions zones or from an encryption zone to unencrypted paths.

The NameNode and HDFS client interact with the Hadoop KMS (or an alternate KMS you configured) through the KeyProvider API operation. The KMS is responsible for storing encryption keys in the backing keystore. Also, Amazon EMR includes the JCE unlimited strength policy, so you can create keys at a desired length.

For more information, see Transparent encryption in HDFS in the Hadoop documentation.

> **Note**
>
> In Amazon EMR, KMS over HTTPS is not enabled by default with Hadoop KMS. For more information about how to enable KMS over HTTPS, see the Hadoop KMS documentation.

Configuring HDFS transparent encryption

You can configure transparent encryption in Amazon EMR by creating keys and adding encryption zones. You can do this in several ways:

- Using the Amazon EMR configuration API operation when you create a cluster
- Using a Hadoop JAR step with command-runner.jar
- Logging in to the master node of the Hadoop cluster and using the `hadoop key` and `hdfs crypto` command line clients
- Using the REST APIs for Hadoop KMS and HDFS

For more information about the REST APIs, see the respective documentation for Hadoop KMS and HDFS.

To create encryption zones and their keys at cluster creation using the CLI

The `hdfs-encryption-zones` classification in the configuration API operation allows you to specify a key name and an encryption zone when you create a cluster. Amazon EMR creates this key in Hadoop KMS on your cluster and configures the encryption zone.

- Create a cluster with the following command.

```
aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge --
instance-count 2 \
--applications Name=App1 Name=App2 --configurations https://s3.amazonaws.com/
mybucket/myfolder/myConfig.json
```

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

`myConfig.json`:

```
[
  {
    "Classification": "hdfs-encryption-zones",
    "Properties": {
      "/myHDFSPath1": "path1_key",
      "/myHDFSPath2": "path2_key"
    }
  }
]
```

To create encryption zones and their keys manually on the master node

1. Launch your cluster using an Amazon EMR release greater than 4.1.0.
2. Connect to the master node of the cluster with SSH.
3. Create a key within Hadoop KMS.

```
$ hadoop key create path2_key
path2_key has been successfully created with options Options{cipher='AES/CTR/
NoPadding', bitLength=256, description='null', attributes=null}.
KMSClientProvider[http://ip-x-x-x-x.ec2.internal:16000/kms/v1/] has been updated.
```

> **Important**
>
> Hadoop KMS requires your key names to be lowercase. If you use a key that has uppercase characters, then your cluster will fail during launch.

4. Create the encryption zone path in HDFS.

```
$ hadoop fs -mkdir /myHDFSPath2
```

5. Make the HDFS path an encryption zone using the key that you created.

```
$ hdfs crypto -createZone -keyName path2_key -path /myHDFSPath2
Added encryption zone /myHDFSPath2
```

To create encryption zones and their keys manually using the AWS CLI

- Add steps to create the KMS keys and encryption zones manually with the following command.

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF --steps Type=CUSTOM_JAR,Name="Create
 First Hadoop KMS Key",Jar="command-runner.jar",ActionOnFailure=CONTINUE,Args=[/
bin/bash,-c,"\"hadoop key create path1_key\""] \
Type=CUSTOM_JAR,Name="Create First Hadoop HDFS Path",Jar="command-
runner.jar",ActionOnFailure=CONTINUE,Args=[/bin/bash,-c,"\"hadoop fs -mkdir /
myHDFSPath1\""] \
Type=CUSTOM_JAR,Name="Create First Encryption Zone",Jar="command-
runner.jar",ActionOnFailure=CONTINUE,Args=[/bin/bash,-c,"\"hdfs crypto -createZone
 -keyName path1_key -path /myHDFSPath1\""] \
Type=CUSTOM_JAR,Name="Create Second Hadoop KMS Key",Jar="command-
runner.jar",ActionOnFailure=CONTINUE,Args=[/bin/bash,-c,"\"hadoop key create
 path2_key\""] \
Type=CUSTOM_JAR,Name="Create Second Hadoop HDFS Path",Jar="command-
runner.jar",ActionOnFailure=CONTINUE,Args=[/bin/bash,-c,"\"hadoop fs -mkdir /
myHDFSPath2\""] \
Type=CUSTOM_JAR,Name="Create Second Encryption Zone",Jar="command-
runner.jar",ActionOnFailure=CONTINUE,Args=[/bin/bash,-c,"\"hdfs crypto -createZone
 -keyName path2_key -path /myHDFSPath2\""]
```

Note

Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

Considerations for HDFS transparent encryption

A best practice is to create an encryption zone for each application where they may write files. Also, you can encrypt all of HDFS by using the hdfs-encryption-zones classification in the configuration API and specify the root path (/) as the encryption zone.

Hadoop key management server

Hadoop KMS is a key management server that provides the ability to implement cryptographic services for Hadoop clusters, and can serve as the key vendor for Transparent encryption in HDFS on Amazon EMR. Hadoop KMS in Amazon EMR is installed and enabled by default when you select the Hadoop application while launching an EMR cluster. The Hadoop KMS does not store the keys itself except in the case of temporary caching. Hadoop KMS acts as a proxy between the key provider and the client trustee to a backing keystore—it is not a keystore. The default keystore that is created for Hadoop KMS is the Java Cryptography Extension KeyStore (JCEKS). The JCE unlimited strength policy is also included, so you can create keys with the desired length. Hadoop KMS also supports a range of ACLs that control access to keys and key operations independently of other client applications such as HDFS. The default key length in Amazon EMR is 256 bit.

To configure Hadoop KMS, use the hadoop-kms-site classification to change settings. To configure ACLs, you use the classification kms-acls.

For more information, see the Hadoop KMS documentation. Hadoop KMS is used in Hadoop HDFS transparent encryption. To learn more about HDFS transparent encryption, see the HDFS transparent encryption topic in the Apache Hadoop documentation.

Note

In Amazon EMR, KMS over HTTPS is not enabled by default with Hadoop KMS. To learn how to enable KMS over HTTPS, see the Hadoop KMS documentation.

Important

Hadoop KMS requires your key names to be lowercase. If you use a key that has uppercase characters, then your cluster will fail during launch.

Configuring Hadoop KMS in Amazon EMR

Using Amazon EMR release version 4.6.0 or later, the `kms-http-port` is 9700 and `kms-admin-port` is 9701.

You can configure Hadoop KMS at cluster creation time using the configuration API for Amazon EMR releases. The following are the configuration object classifications available for Hadoop KMS:

Hadoop KMS configuration classifications

Classification	Filename
hadoop-kms-site	`kms-site.xml`
hadoop-kms-acls	`kms-acls.xml`
hadoop-kms-env	`kms-env.sh`
hadoop-kms-log4j	`kms-log4j.properties`

To set Hadoop KMS ACLs using the CLI

- Create a cluster with Hadoop KMS with ACLs using the following command:

```
aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge --
instance-count 2 \
--applications Name=App1 Name=App2 --configurations https://s3.amazonaws.com/
mybucket/myfolder/myConfig.json
```

Note

Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

`myConfig.json`:

```
[
    {
      "Classification": "hadoop-kms-acls",
      "Properties": {
        "hadoop.kms.blacklist.CREATE": "hdfs,foo,myBannedUser",
        "hadoop.kms.acl.ROLLOVER": "myAllowedUser"
      }
    }
  ]
```

To disable Hadoop KMS cache using the CLI

- Create a cluster with Hadoop KMS `hadoop.kms.cache.enable` set to `false`, using the following command:

```
aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge --
instance-count 2 \
--applications Name=App1 Name=App2 --configurations https://s3.amazonaws.com/
mybucket/myfolder/myConfig.json
```

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

`myConfig.json`:

```
[
    {
      "Classification": "hadoop-kms-site",
      "Properties": {
        "hadoop.kms.cache.enable": "false"
      }
    }
  ]
```

To set environment variables in the kms-env.sh script using the CLI

- Change settings in kms-env.sh via the hadoop-kms-env configuration. Create a cluster with Hadoop KMS using the following command:

```
aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge --
instance-count 2 \
--applications Name=App1 Name=App2 --configurations https://s3.amazonaws.com/
mybucket/myfolder/myConfig.json
```

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

myConfig.json:

```
[
  {
    "Classification": "hadoop-kms-env",
    "Properties": {
    },
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "JAVA_LIBRARY_PATH": "/path/to/files",
          "KMS_SSL_KEYSTORE_FILE": "/non/Default/Path/.keystore",
          "KMS_SSL_KEYSTORE_PASS": "myPass"
        },
        "Configurations": [
        ]
      }
    ]
  }
]
```

For information about configuring Hadoop KMS, see the Hadoop KMS documentation.

HDFS transparent encryption on EMR clusters with multiple master nodes

Apache Ranger KMS is used in an Amazon EMR cluster with multiple primary nodes for transparent encryption in HDFS.

Apache Ranger KMS stores its root key and Encryption Zone (EZ) keys in your Amazon RDS for an Amazon EMR cluster with multiple primary nodes. To enable transparent encryption in HDFS on an Amazon EMR cluster with multiple primary nodes, you must provide the following configurations.

- Amazon RDS or your own MySQL server connection URL to store the Ranger KMS root key and EZ key
- User name and password for MySQL
- Password for Ranger KMS root key
- Certificate Authority (CA) PEM file for SSL connection to MySQL server

You can provide these configurations by using `ranger-kms-dbks-site` classification and `ranger-kms-db-ca` classification, as the following example demonstrates.

```
[
  {
    "Classification": "ranger-kms-dbks-site",
    "Properties": {
      "ranger.ks.jpa.jdbc.url": "jdbc:log4jdbc:mysql://mysql-host-url.xx-
xxx-1.xxx.amazonaws.com:3306/rangerkms",
      "ranger.ks.jpa.jdbc.user": "mysql-user-name",
      "ranger.ks.jpa.jdbc.password": "mysql-password",
      "ranger.db.encrypt.key.password": "password-for-encrypting-a-master-key"
    }
  },
  {
    "Classification": "ranger-kms-db-ca",
    "Properties": {
      "ranger.kms.trust.ca.file.s3.url": "s3://rds-downloads/rds-ca-2019-root.pem"
    }
     }
]
```

The following are configuration object classifications for Apache Ranger KMS.

Hadoop KMS configuration classifications

Classification	Description
ranger-kms-dbks-site	Change values in dbks-site.xml file of Ranger KMS.
ranger-kms-site	Change values in ranger-kms-site.xml file of Ranger KMS.
ranger-kms-env	Change values in the Ranger KMS environme nt.
ranger-kms-log4j	Change values in kms-log4j.properties file of Ranger KMS.
ranger-kms-db-ca	Change values for CA file on S3 for MySQL SSL connection with Ranger KMS.

Considerations

- It is highly recommended that you encrypt your Amazon RDS instance to improve security. For more information, see Overview of encrypting Amazon RDS resources.
- It is highly recommended that you use separate MySQL database for each Amazon EMR cluster with multiple primary nodes for high security bar.
- To configure transparent encryption in HDFS on an Amazon EMR cluster with multiple primary nodes, you must specify the `hdfs-encryption-zones` classification while creating the cluster. Otherwise, Ranger KMS will not be configured or started. Reconfiguring `hdfs-encryption-zones` classification or any of the Hadoop KMS configuration classifications on a running cluster is not supported on Amazon EMR cluster with multiple primary nodes.

Create or run a Hadoop application

Topics

- Process data with a custom JAR

Build binaries using Amazon EMR

You can use Amazon EMR as a build environment to compile programs for use in your cluster. Programs that you use with Amazon EMR must be compiled on a system running the same version of Linux used by Amazon EMR. For a 32-bit version, you should have compiled on a 32-bit machine or with 32-bit cross compilation options turned on. For a 64-bit version, you need to have compiled on a 64-bit machine or with 64-bit cross compilation options turned. For more information about EC2 instance versions, see Plan and configure EC2 instances in the *Amazon EMR Management Guide*. Supported programming languages include C++, Python, and C#.

The following table outlines the steps involved to build and test your application using Amazon EMR.

Process for building a module

1	Connect to the master node of your cluster.
2	Copy source files to the master node.
3	Build binaries with any necessary optimizations.
4	Copy binaries from the master node to Amazon S3.

The details for each of these steps are covered in the sections that follow.

To connect to the master node of the cluster

- Follow the instructions at Connect to the master node using SSH in the *Amazon EMR Management Guide*.

To copy source files to the master node

1. Put your source files in an Amazon S3 bucket. To learn how to create buckets and how to move data into Amazon S3, see the Amazon Simple Storage Service User Guide.
2. Create a folder on your Hadoop cluster for your source files by entering a command similar to the following:

```
mkdir SourceFiles
```

3. Copy your source files from Amazon S3 to the master node by typing a command similar to the following:

```
hadoop fs -get s3://mybucket/SourceFiles SourceFiles
```

Build binaries with any necessary optimizations

How you build your binaries depends on many factors. Follow the instructions for your specific build tools to setup and configure your environment. You can use Hadoop system specification commands to obtain cluster information to determine how to install your build environment.

To identify system specifications

- Use the following commands to verify the architecture you are using to build your binaries.

  a. To view the version of Debian, enter the following command:

  ```
  master$ cat /etc/issue
  ```

  The output looks similar to the following.

  ```
  Debian GNU/Linux 5.0
  ```

  b. To view the public DNS name and processor size, enter the following command:

  ```
  master$ uname -a
  ```

  The output looks similar to the following.

  ```
  Linux domU-12-31-39-17-29-39.compute-1.internal 2.6.21.7-2.fc8xen #1 SMP Fri
   Feb 15 12:34:28 EST 2008 x86_64 GNU/Linux
  ```

  c. To view the processor speed, enter the following command:

  ```
  master$ cat /proc/cpuinfo
  ```

The output looks similar to the following.

```
processor : 0
vendor_id : GenuineIntel
model name : Intel(R) Xeon(R) CPU E5430 @ 2.66GHz
flags : fpu tsc msr pae mce cx8 apic mca cmov pat pse36 clflush dts acpi mmx
 fxsr sse sse2 ss ht tm syscall nx lm constant_tsc pni monitor ds_cpl vmx est
 tm2 ssse3 cx16 xtpr cda lahf_lm
...
```

Once your binaries are built, you can copy the files to Amazon S3.

To copy binaries from the master node to Amazon S3

- Type the following command to copy the binaries to your Amazon S3 bucket:

```
hadoop fs -put BinaryFiles s3://mybucket/BinaryDestination
```

Process data with streaming

Hadoop streaming is a utility that comes with Hadoop that enables you to develop MapReduce executables in languages other than Java. Streaming is implemented in the form of a JAR file, so you can run it from the Amazon EMR API or command line just like a standard JAR file.

This section describes how to use streaming with Amazon EMR.

Note

Apache Hadoop streaming is an independent tool. As such, all of its functions and parameters are not described here. For more information about Hadoop streaming, go to http://hadoop.apache.org/docs/stable/hadoop-streaming/HadoopStreaming.html.

Using the Hadoop streaming utility

This section describes how use to Hadoop's streaming utility.

Hadoop process

1	Write your mapper and reducer executable in the programming language of your choice. Follow the directions in Hadoop's documentation to write your streaming executables. The programs should read their input from standard input and output data through standard output. By default, each line of input/output represents a record and the first tab on each line is used as a separator between the key and value.
2	Test your executables locally and upload them to Amazon S3.
3	Use the Amazon EMR command line interface or Amazon EMR console to run your application.

Each mapper script launches as a separate process in the cluster. Each reducer executable turns the output of the mapper executable into the data output by the job flow.

The `input`, `output`, `mapper`, and `reducer` parameters are required by most streaming applications. The following table describes these and other, optional parameters.

Parameter	Description	Required
-input	Location on Amazon S3 of the input data. Type: String Default: None Constraint: URI. If no protocol is specified then it uses the clust er's default file system.	Yes
-output	Location on Amazon S3 where Amazon EMR uploads the processed data. Type: String Default: None	Yes

Parameter	Description	Required
	Constraint: URI Default: If a location is not specified, Amazon EMR uploads the data to the location specified by `input`.	
-mapper	Name of the mapper executable. Type: String Default: None	Yes
-reducer	Name of the reducer executable. Type: String Default: None	Yes
-cacheFile	An Amazon S3 location containing files for Hadoop to copy into your local working directory (primarily to improve performance). Type: String Default: None Constraints: [URI]#[symlink name to create in working directory]	No
-cacheArchive	JAR file to extract into the working directory Type: String Default: None Constraints: [URI]#[symlink directory name to create in working directory	No

Parameter	Description	Required
-combiner	Combines results Type: String Default: None Constraints: Java class name	No

The following code sample is a mapper executable written in Python. This script is part of the WordCount sample application.

```
#!/usr/bin/python
import sys

def main(argv):
  line = sys.stdin.readline()
  try:
    while line:
      line = line.rstrip()
      words = line.split()
      for word in words:
        print "LongValueSum:" + word + "\t" + "1"
      line = sys.stdin.readline()
  except "end of file":
    return None
if __name__ == "__main__":
  main(sys.argv)
```

Submit a streaming step

This section covers the basics of submitting a streaming step to a cluster. A streaming application reads input from standard input and then runs a script or executable (called a mapper) against each input. The result from each of the inputs is saved locally, typically on a Hadoop Distributed File System (HDFS) partition. After all the input is processed by the mapper, a second script or executable (called a reducer) processes the mapper results. The results from the reducer are sent to standard output. You can chain together a series of streaming steps, where the output of one step becomes the input of another step.

The mapper and the reducer can each be referenced as a file or you can supply a Java class. You can implement the mapper and reducer in any of the supported languages, including Ruby, Perl, Python, PHP, or Bash.

Submit a streaming step using the console

This example describes how to use the Amazon EMR console to submit a streaming step to a running cluster.

To submit a streaming step

1. Open the Amazon EMR console at https://console.aws.amazon.com/emr.
2. In the **Cluster List**, select the name of your cluster.
3. Scroll to the **Steps** section and expand it, then choose **Add step**.
4. In the **Add Step** dialog box:
   - For **Step type**, choose **Streaming program**.
   - For **Name**, accept the default name (Streaming program) or type a new name.
   - For **Mapper**, type or browse to the location of your mapper class in Hadoop, or an S3 bucket where the mapper executable, such as a Python program, resides. The path value must be in the form *`BucketName`*/*`path`*/*`MapperExecutable`*.
   - For **Reducer**, type or browse to the location of your reducer class in Hadoop, or an S3 bucket where the reducer executable, such as a Python program, resides. The path value must be in the form *`BucketName`*/*`path`*/*`MapperExecutable`*. Amazon EMR supports the special *aggregate* keyword. For more information, go to the Aggregate library supplied by Hadoop.
   - For **Input S3 location**, type or browse to the location of your input data.
   - For **Output S3 location**, type or browse to the name of your Amazon S3 output bucket.
   - For **Arguments**, leave the field blank.
   - For **Action on failure**, accept the default option **(Continue**).
5. Choose **Add**. The step appears in the console with a status of Pending.
6. The status of the step changes from Pending to Running to Completed as the step runs. To update the status, choose the **Refresh** icon above the Actions column.

AWS CLI

These examples demonstrate how to use the AWS CLI to create a cluster and submit a Streaming step.

To create a cluster and submit a streaming step using the AWS CLI

- To create a cluster and submit a streaming step using the AWS CLI, type the following command and replace *myKey* with the name of your EC2 key pair. Note that your argument for `--files` should be the Amazon S3 path to your script's location, and the arguments for `-mapper` and `-reducer` should be the names of the respective script files.

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.2.0 --
applications Name=Hue Name=Hive Name=Pig --use-default-roles \
--ec2-attributes KeyName=myKey --instance-type m5.xlarge --instance-count 3 \
--steps Type=STREAMING,Name="Streaming Program",ActionOnFailure=CONTINUE,Args=[--
files,pathtoscripts,-mapper,mapperscript,-reducer,reducerscript,aggregate,-
input,pathtoinputdata,-output,pathtooutputbucket]
```

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

When you specify the instance count without using the `--instance-groups` parameter, a single master node is launched, and the remaining instances are launched as core nodes. All nodes use the instance type specified in the command.

> **Note**
>
> If you have not previously created the default Amazon EMR service role and EC2 instance profile, type aws `emr create-default-roles` to create them before typing the `create-cluster` subcommand.

For more information on using Amazon EMR commands in the AWS CLI, see https://docs.aws.amazon.com/cli/latest/reference/emr.

Process data with a custom JAR

A custom JAR runs a compiled Java program that you can upload to Amazon S3. You should compile the program against the version of Hadoop you want to launch, and submit a `CUSTOM_JAR` step to your Amazon EMR cluster. For more information about how to compile a JAR file, see Build binaries using Amazon EMR.

For more information about building a Hadoop MapReduce application, see the MapReduce Tutorial in the Apache Hadoop documentation.

Topics

Submit a custom JAR step

A custom JAR runs a compiled Java program that you can upload to Amazon S3. You should compile the program against the version of Hadoop you want to launch, and submit a `CUSTOM_JAR` step to your Amazon EMR cluster. For more information about how to compile a JAR file, see Build binaries using Amazon EMR.

For more information about building a Hadoop MapReduce application, see the MapReduce Tutorial in the Apache Hadoop documentation.

This section covers the basics of submitting a custom JAR step in Amazon EMR. Submitting a custom JAR step enables you to write a script to process your data with the Java programming language.

Submit a custom JAR step with the console

This example describes how to use the Amazon EMR console to submit a custom JAR step to a running cluster.

To submit a custom JAR step with the console

1. Open the Amazon EMR console at https://console.aws.amazon.com/emr.
2. In the **Cluster List**, select the name of your cluster.
3. Scroll to the **Steps** section and expand it, then choose **Add step**.
4. In the **Add Step** dialog:

- For **Step type**, choose **Custom JAR**.
- For **Name**, accept the default name (Custom JAR) or type a new name.
- For **JAR S3 location**, type or browse to the location of your JAR file. JAR location maybe a path into S3 or a fully qualified java class in the classpath..
- For **Arguments**, type any required arguments as space-separated strings or leave the field blank.
- For **Action on failure**, accept the default option (**Continue**).

5. Choose **Add**. The step appears in the console with a status of Pending.
6. The status of the step changes from Pending to Running to Completed as the step runs. To update the status, choose the **Refresh** icon above the Actions column.

Launching a cluster and submitting a custom JAR step with the AWS CLI

To launch a cluster and submit a custom JAR step with the AWS CLI

To launch a cluster and submit a custom JAR step with the AWS CLI, type the `create-cluster` subcommand with the `--steps` parameter.

- To launch a cluster and submit a custom JAR step, type the following command, replace *myKey* with the name of your EC2 key pair, and replace *mybucket* with your bucket name.

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.2.0 \
--applications Name=Hue Name=Hive Name=Pig --use-default-roles \
--ec2-attributes KeyName=myKey --instance-type m5.xlarge --instance-count 3 \
--steps Type=CUSTOM_JAR,Name="Custom JAR
 Step",ActionOnFailure=CONTINUE,Jar=pathtojarfile,Args=["pathtoinputdata","pathtooutputbuck
```

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

When you specify the instance count without the `--instance-groups` parameter, a single primary node launches, and the remaining instances launch as core nodes. All nodes use the instance type that you specify in the command.

> **Note**
>
> If you have not previously created the default Amazon EMR service role and EC2 instance profile, type `aws emr create-default-roles` to create them before typing the `create-cluster` subcommand.

For more information on using Amazon EMR commands in the AWS CLI, see https://docs.aws.amazon.com/cli/latest/reference/emr.

Third-party dependencies

Sometimes it may be necessary to include in the MapReduce classpath JARs for use with your program. You have two options for doing this:

- Include the `--libjars s3://URI_to_JAR` in the step options for the procedure in Launching a cluster and submitting a custom JAR step with the AWS CLI.
- Launch the cluster with a modified `mapreduce.application.classpath` setting in `mapred-site.xml`. Use the `mapred-site` configuration classification. To create the cluster with the step using AWS CLI, this would look like the following:

```
aws emr create-cluster --release-label emr-7.2.0 \
--applications Name=Hue Name=Hive Name=Pig --use-default-roles \
--instance-type m5.xlarge --instance-count 2  --ec2-attributes KeyName=myKey \
--steps Type=CUSTOM_JAR,Name="Custom JAR
 Step",ActionOnFailure=CONTINUE,Jar=pathtojarfile,Args=["pathtoinputdata","pathtooutputbucket
 \
--configurations https://s3.amazonaws.com/mybucket/myfolder/myConfig.json
```

`myConfig.json`:

```
[
    {
      "Classification": "mapred-site",
      "Properties": {
        "mapreduce.application.classpath": "path1,path2"
      }
    }
```

```
  ]
```

The comma-separated list of paths should be appended to the JVM classpath for each task.

Read restored objects

With Amazon EMR release 7.2.0 and higher, you can read restored Glacier objects from the S3 location of the table with the `S3A` protocol. Engines in previous releases don't distinguish between Glacier and Glacier Deep Archive files, which means you would get an `AmazonS3Exception` if you tried to access an in-progress Glacier file with `S3A`. This read operation ignores archived Glacier files if they still in progress of being restored. To enable this behavior, use the setting `fs.s3a.glacier.read.restored.objects`. This setting can be three values:

- **READ_ALL** – this value indicates that Amazon EMR shouldn't account for the storage classes retrieved from Amazon S3. This is the default behavior.
- **SKIP_ALL_GLACIER** – this value indicates that Amazon EMR should ignore any S3 objects that are tagged with the Glacier storage class and retrieve all other objects. This is the default behavior for Amazon Athena with respect to Glacier objects.
- **READ_RESTORED_GLACIER_OBJECTS** – this value indicates that Amazon EMR should check the restored status of the Glacier object. If Amazon EMR can restore the object, you can read them like a normal S3 object. Otherwise, Amazon EMR ignores the object from S3 Glacier

Examples

Spark

To read restored objects when you use Spark, use the following configuration:

```
--conf spark.hadoop.fs.s3a.glacier.read.restored.objects=<value>
```

If you use spark-sql, use the following configuration instead;

```
spark-sql --conf spark.hadoop.fs.s3a.glacier.read.restored.objects=<value>
```

Flink

If you use Flink, you can set the configuration in the `flink-conf.yaml` file

```
fs.s3a.glacier.read.restored.objects: <value>
```

You can also set the `flink-conf` classification:

```
[
    {
      "Classification": "flink-conf",
      "Properties": {
        "fs.s3a.glacier.read.restored.objects":"<value>"
      }
    }
]
```

Hive

If you use Hive, set the configuration in the `hive-site.xml` file.

```
<property>
    <name>fs.s3a.glacier.read.restored.objects</name>
    <value><value></value>
</property>
```

You can also use the Hive CLI to set the property `--hiveconf`:

```
hive --hiveconf fs.s3a.glacier.read.restored.objects=<value>
```

Considerations

When you read restored objects from S3 Glacier, note the following considerations:

- You can only read restored objects if you use the `S3A` scheme or the `S3AFileSystem` to access the data.
- When you read a restored Glacier object, Amazon EMR doesn't restore the object itself. To do so, you must use the AWS CLI or the AWS SDK.

Turn on non-uniform memory access awareness for YARN containers

With Amazon EMR versions 6.x and later, you can use non-uniform memory access (NUMA) for multiprocessing your data on clusters. NUMA is a computer memory design pattern where the processor can access its own local memory faster than memory on another processor or shared between processors. YARN containers have better performance with NUMA because they can bind to a specific NUMA node that serves all subsequent memory allocations. This reduces the amount of times that your cluster has to access remote memory.

You can turn on NUMA support for YARN container when the worker node machine is a multi-NUMA node. To confirm if a worker node is a single-NUMA or multi-NUMA node, run the following command.

```
lscpu | grep -i numa
NUMA node(s): 2
```

In general, instances that are larger than 12x have two NUMA nodes. This does not apply to metal instances.

To turn on NUMA awareness for YARN containers

1. Use the following `yarn-site` configuration in your Amazon EMR 6.x cluster.

```
 [
    {
        "classification":"yarn-site",
            "properties":{
                "yarn.nodemanager.linux-container-executor.nonsecure-mode.local-
user":"yarn",
                "yarn.nodemanager.linux-container-executor.group":"yarn",
                "yarn.nodemanager.container-
executor.class":"org.apache.hadoop.yarn.server.nodemanager.LinuxContainerExecutor",
                "yarn.nodemanager.numa-awareness.enabled":"true",
                "yarn.nodemanager.numa-awareness.numactl.cmd":"/usr/bin/numactl",
                "yarn.nodemanager.numa-awareness.read-topology":"true"
            },
        "configurations":[]
     }
 ]
```

2. Provide the following bootstrap action in your cluster.

```
#!/bin/bash

sudo yum -y install numactl
echo 1 | sudo tee /proc/sys/kernel/numa_balancing

echo "banned.users=mapred,bin,hdfs" >> /etc/hadoop/conf/container-executor.cfg
rm -rf /var/log/hadoop-yarn/
sudo chown -R yarn:hadoop /var/log/hadoop-yarn/
sudo chmod 755 -R /var/log/hadoop-yarn/

sudo chmod 6050 /etc/hadoop/conf/container-executor.cfg

mkdir /mnt/yarn && sudo chmod 755 -R /mnt/yarn && sudo chown -R yarn:hadoop /mnt/
yarn
mkdir /mnt1/yarn && sudo chmod 755 -R /mnt1/yarn && sudo chown -R yarn:hadoop /
mnt1/yarn
mkdir /mnt2/yarn && sudo chmod 755 -R /mnt2/yarn && sudo chown -R yarn:hadoop /
mnt2/yarn
```

3. Every container must be aware of NUMA. You can notify the Java virtual machine (JVM) in each container with a NUMA flag. For example, to notify the JVM to use NUMA in a MapReduce job, add the following properties in `mapred-site.xml`.

```
<property>
    <name>mapreduce.reduce.java.opts</name>
    <value>-XX:+UseNUMA</value>
</property>
<property>
    <name>mapreduce.map.java.opts</name>
    <value>-XX:+UseNUMA</value>
</property>
```

4. To verify that you turned NUMA on, search any of the NodeManager log files with the following command.

```
grep "NUMA resources allocation is enabled," *
```

To verify that NodeManager has assigned NUMA node resources to a container, search the NodeManager log with the following command, replacing *<container_id>* with your own container ID.

```
grep "NUMA node" | grep <container_id>
```

Hadoop version history

The following table lists the version of Hadoop included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in Amazon EMR 7.x release versions, Amazon EMR 6.x release versions, or Amazon EMR 5.x release versions.

Hadoop version information

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-7.2.0	3.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.36.2	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server,

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
		hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-7.1.0	3.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-7.0.0	3.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-6.15.0	3.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.14.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-6.13.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.12.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-6.11.1	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.11.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-6.10.1	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.10.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-6.9.1	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.9.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-6.8.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.8.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-6.7.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.36.1	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.36.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.6.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.35.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.5.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-6.4.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.3.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-6.3.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.2.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-6.2.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.1.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-6.1.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.0.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-6.0.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.34.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.33.1	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.33.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.32.1	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.32.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.31.1	2.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.31.0	2.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.30.2	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.30.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.30.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.29.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.28.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.28.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.27.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.27.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.26.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.25.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.24.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.24.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.23.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.23.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.22.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.21.2	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.21.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.21.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.20.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.20.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.19.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.19.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.18.1	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.18.0	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.17.2	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.17.1	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.17.0	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.16.1	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.16.0	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.15.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.15.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.14.2	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.14.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.14.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.13.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.13.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.12.3	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.12.2	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.12.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.12.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.11.4	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.11.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.11.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.11.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.11.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.10.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.10.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.9.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.9.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.8.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.8.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.8.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.8.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.7.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.7.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.6.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.6.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.5.4	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.5.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.5.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.5.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.5.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.4.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.4.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.3.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.3.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.3.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.2.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.2.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.2.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.2.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.1.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.1.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.0.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-5.0.2	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.0.1	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.0.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-4.9.6	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.9.5	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.9.4	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-4.9.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.9.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.9.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-4.8.5	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.8.4	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.8.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-4.8.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.8.1	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.8.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-4.7.4	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.7.3	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.7.2	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-4.7.1	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.7.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.6.1	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-4.6.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.5.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.4.0	2.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-4.3.0	2.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.2.0	2.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.1.0	2.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release label	Hadoop Version	Components installed with Hadoop
emr-4.0.0	2.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Hadoop release notes by version

Topics

Amazon EMR 7.2.0 - Hadoop release notes

Amazon EMR 7.2.0 - Hadoop changes

Type	Description
New Feature	HADOOP-18850: S3A: Enable dual-layer server-side encryption with AWS KMS keys
New Feature	Support S3 Glacier read restored objects
Improvement	Dynamically Scaling S3A maximum connections based upon the cores when running Spark Jobs

Type	Description
Improvement	HADOOP-18797: Support Concurrent Writes With S3A Magic Committer
Improvement	HADOOP-19047: Support InMemory Tracking Of S3A Magic Commits
Improvement	Add "Export", "Publish" & "AbortPublish" method in MagicV2 Committer
Improvement	Implement ReadFullyIntoBuffers FileSystem API in S3A
Bug Fix	HADOOP-18793: S3A StagingCommitter does not clean up staging-uploads directory
Upgrade	HADOOP-19008: S3A: update aws-sdk version to 2.21.41
Upgrade	HADOOP-18613: Upgrade ZooKeeper to version 3.8.3

Amazon EMR 7.2.0 - Hadoop features

See the following list for new Hadoop features in Amazon EMR 7.2.0.

- S3 glacier support to read restored objects – With Amazon EMR release 7.2.0 and higher, you can read restored Glacier objects from the S3 location of the table with the S3A protocol. This read operation ignores archived Glacier files if they still in progress of being restored.
- Support concurrent writes with the S3A magic committer – Amazon EMR 7.2.0 introduces support for concurrent writes with the S3A magic committer. This feature was added to the OSS Hadoop community as part of HADOOP-18797.

Amazon EMR 6.6.0 - Hadoop release notes

Amazon EMR 6.6.0 - Hadoop changes

Type	Description
Bug	Fixed duplicated records when reading BZip2 text files.
Backport	HADOOP-18136: Verify `FileUtils.unTar()` handling of missing .tar files
Backport	HADOOP-17627: Backport to branch-3.2 HADOOP-17371, HADOOP-17621, HADOOP-17625 to update Jetty to 9.4.39
Backport	HADOOP-17655: Upgrade Jetty to 9.4.40
Backport	HADOOP-17796: Upgrade jetty version to 9.4.43
Backport	HADOOP-17661: mvn versions:set fails to parse pom.xml
Backport	HADOOP-17236: Bump up snakeyaml to 1.26 to mitigate CVE-2017-18640
Backport	HADOOP-16717: Remove GenericsUtil isLog4jLogger dependency on Log4jLoggerAdapter
Backport	HADOOP-17633: Bump json-smart to 2.4.2 and nimbus-jose-jwt to 9.8 due to CVEs
Backport	HADOOP-17844: Upgrade JSON smart to 2.4.7
Backport	HADOOP-17972: Backport HADOOP-17683 (Update commons-io to 2.8.0) for branch-3.2

Type	Description
Backport	HADOOP-16555: Update commons-compress to 1.19
Backport	HADOOP-17370: Upgrade commons-c ompress to 1.21
Backport	HADOOP-17096: Fix ZStandardCompressor input buffer offset
Backport	HADOOP-17112: Whitespace not allowed in paths when saving files to s3a via committer
Backport	HADOOP-13500: Synchronizing iteration of Configuration properties object
Backport	HDFS-14099: Unknown frame descriptor when decompressing multiple frames in ZStandardDecompressor
Backport	HDFS-16410: Insecure Xml parsing in Offli neEditsXmlLoader
Backport	HDFS-14498: LeaseManager can loop forever on the file for which create has failed
Backport	HDFS-15290: NPE in HttpServer during NameNode startup
Backport	HDFS-15293: Relax the condition for accepting a fsimage when receiving a checkpoint
Backport	HDFS-12979: StandbyNode should upload FsImage to ObserverNode after checkpointing
Backport	YARN-10538: Add recommissioning nodes to the list of updated nodes returned to the AM

Type	Description
Backport	YARN-10472: Backport YARN-10314 (YarnClient throws NoClassDefFoundError for WebSocketException with only shaded client jars) to branch-3.2
Backport	YARN-9968: Public Localizer is exiting in `NodeManager` due to `NullPointerException`
Backport	YARN-10651: `CapacityScheduler` crashed with `NPE` in `AbstractYarnScheduler.updateNodeResource()`
Backport	YARN-9339: Apps pending metric incorrect after moving app to a new queue
Backport	YARN-10438: Handle null `containerId` in `ClientRMService#getContainerReport()`
Backport	YARN-7266: ATS 1.5 fails to start if `RollingLevelDb` files are corrupt or missing
Backport	YARN-9063: ATS 1.5 fails to start if `RollingLevelDb` files are corrupt or missing
Backport	YARN-9848: Revert YARN-4946 (RM should not consider an application as `COMPLETED` when log aggregation is not in a terminal state).

Apache HBase

HBase is an open source, non-relational, distributed database developed as part of the Apache Software Foundation's Hadoop project. HBase runs on top of Hadoop Distributed File System (HDFS) to provide non-relational database capabilities for the Hadoop ecosystem. HBase is included with Amazon EMR release version 4.6.0 and later.

HBase works seamlessly with Hadoop, sharing its file system and serving as a direct input and output to the MapReduce framework and execution engine. HBase also integrates with Apache Hive, enabling SQL-like queries over HBase tables, joins with Hive-based tables, and support for Java Database Connectivity (JDBC). For more information about HBase, see Apache HBase and HBase documentation on the Apache website. For an example of how to use HBase with Hive, see the AWS Big Data Blog post Combine NoSQL and massively parallel analytics using Apache HBase and Apache Hive on Amazon EMR.

With HBase on Amazon EMR, you can also back up your HBase data directly to Amazon Simple Storage Service (Amazon S3), and restore from a previously created backup when launching an HBase cluster. Amazon EMR offers additional options to integrate with Amazon S3 for data persistence and disaster recovery.

- **HBase on Amazon S3** - With Amazon EMR version 5.2.0 and later, you can use HBase on Amazon S3 to store a cluster's HBase root directory and metadata directly to Amazon S3. You can subsequently start a new cluster, pointing it to the root directory location in Amazon S3. Only one cluster at a time can use the HBase location in Amazon S3, with the exception of a read-replica cluster. For more information, see HBase on Amazon S3 (Amazon S3 storage mode).
- **HBase read-replicas** - Amazon EMR version 5.7.0 and later with HBase on Amazon S3 supports read-replica clusters. A read-replica cluster provides read-only access to a primary cluster's store files and metadata for read-only operations. For more information, see Using a read-replica cluster.
- **HBase Snapshots** - As an alternative to HBase on Amazon S3, with EMR version 4.0 and later you can create snapshots of your HBase data directly to Amazon S3 and then recover data using the snapshots. For more information, see Using HBase snapshots.

Important

For Amazon EMR HBase cluster scaling, we do not recommend using managed scaling or scaling with custom policies with HBase clusters.

The following table lists the version of HBase included in the latest release of the Amazon EMR 7.x series, along with the components that Amazon EMR installs with HBase.

For the version of components installed with HBase in this release, see Release 7.2.0 Component Versions.

HBase version information for emr-7.2.0

Amazon EMR Release Label	HBase Version	Components Installed With HBase
emr-7.2.0	HBase 2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

The following table lists the version of HBase included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with HBase.

For the version of components installed with HBase in this release, see Release 6.15.0 Component Versions.

HBase version information for emr-6.15.0

Amazon EMR Release Label	HBase Version	Components Installed With HBase
emr-6.15.0	HBase 2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Note

Apache HBase HBCK2 is a separate operational tool for repairing HBase regions and system tables. In Amazon EMR version 6.1.0 and later, the hbase-hbck2.jar is provided in `/usr/lib/hbase-operator-tools/` on the primary node. For more information about how to build and use the tool, see HBase HBCK2.

The following table lists the version of HBase included in the latest release of the Amazon EMR 5.x series, along with the components that Amazon EMR installs with HBase.

For the version of components installed with HBase in this release, see Release 5.36.2 Component Versions.

HBase version information for emr-5.36.2

Amazon EMR Release Label	HBase Version	Components Installed With HBase
emr-5.36.2	HBase 1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Topics

- Creating a cluster with HBase
- HBase on Amazon S3 (Amazon S3 storage mode)
- Write-ahead logs (WAL) for Amazon EMR
- Using the HBase shell
- Access HBase tables with Hive
- Using HBase snapshots
- Configure HBase
- View the HBase user interface
- View HBase log files

- Monitor HBase with Ganglia
- Migrating from previous HBase versions
- HBase release history

Creating a cluster with HBase

The procedures in this section cover the basics of launching a cluster using the AWS Management Console and the AWS CLI. For detailed information about how to plan, configure, and launch Amazon EMR clusters, see Plan and configure clusters in the *Amazon EMR Management Guide*.

Creating a cluster with HBase using the console

For quick steps to launch clusters with the console, see Getting started with Amazon EMR in the *Amazon EMR Management Guide*.

To launch a cluster with HBase installed using the console

1. Open the Amazon EMR console at https://console.aws.amazon.com/emr.
2. Choose **Create cluster** and **Go to advanced options**.
3. For **Software Configuration**, choose an **Amazon Release Version** of 4.6.0 or later (we recommend the latest version). Choose **HBase** and other applications as desired.
4. With Amazon EMR version 5.2.0 and later, under **HBase Storage Settings**, select **HDFS** or **S3**. For more information, see HBase on Amazon S3 (Amazon S3 storage mode).
5. Select other options as necessary and then choose **Create cluster**.

Creating a cluster with HBase using the AWS CLI

Use the following command to create a cluster with HBase installed:

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.2.0 \
--applications Name=HBase --use-default-roles --ec2-attributes KeyName=myKey \
--instance-type m5.xlarge --instance-count 3
```

Note

Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

If you use HBase on Amazon S3, specify the `--configurations` option with a reference to a JSON configuration object. The configuration object must contain an `hbase-site` classification that specifies the location in Amazon S3 where HBase data is stored using the `hbase.rootdir` property. It also must contain an `hbase` classification, which specifies `s3` using the `hbase.emr.storageMode` property. The following example demonstrates a JSON snippet with these configuration settings.

```
[
    {
        "Classification": "hbase-site",
        "Properties": {
            "hbase.rootdir": "s3://MyBucket/MyHBaseStore"
        }
    },
    {
        "Classification": "hbase",
        "Properties": {
            "hbase.emr.storageMode": "s3"
        }
    }
]
```

For more information about HBase on Amazon S3, see HBase on Amazon S3 (Amazon S3 storage mode). For more information about classifications, see Configure applications.

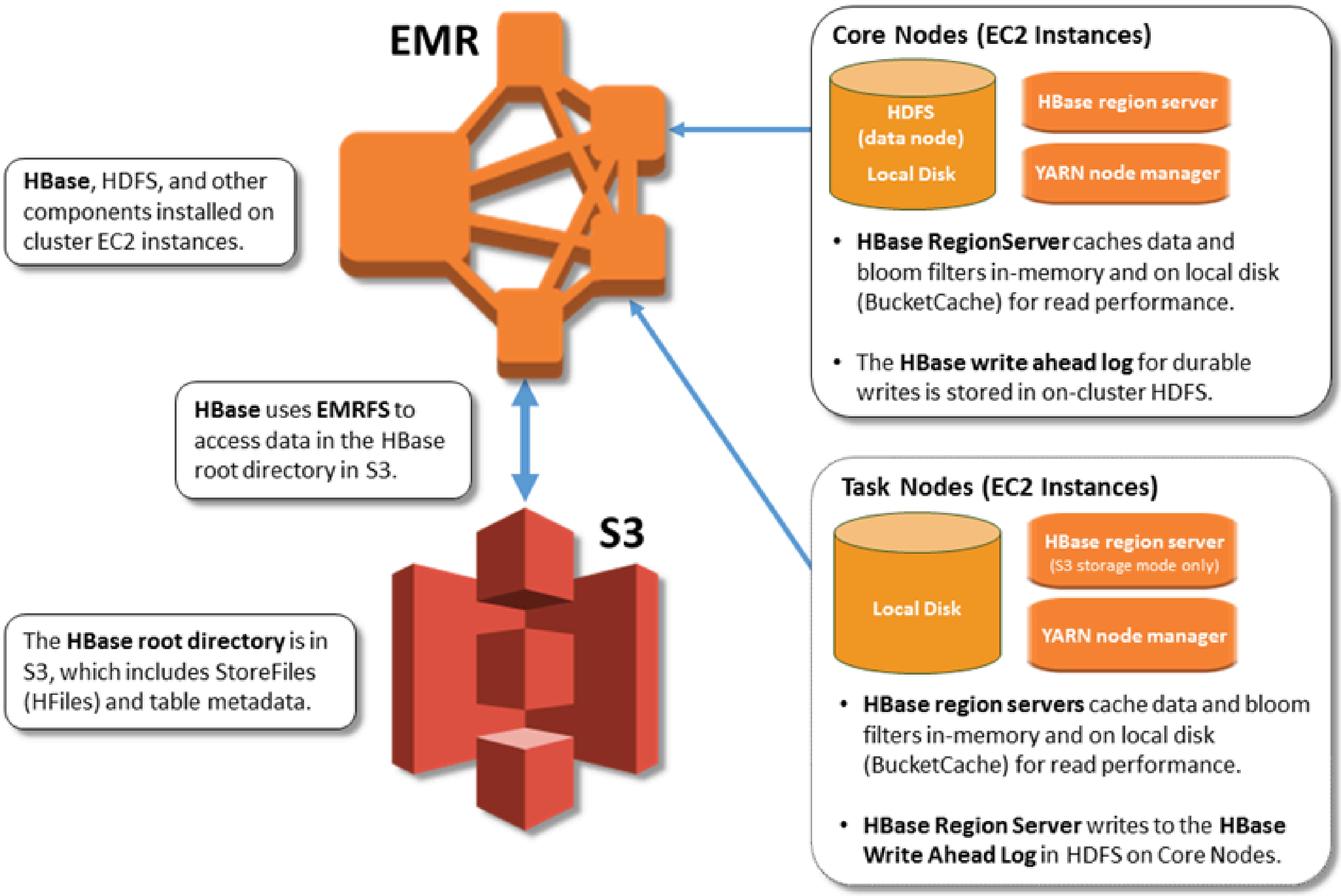
HBase on Amazon S3 (Amazon S3 storage mode)

When you run HBase on Amazon EMR version 5.2.0 or later, you can enable HBase on Amazon S3, which offers the following advantages:

- The HBase root directory is stored in Amazon S3, including HBase store files and table metadata. This data is persistent outside of the cluster, available across Amazon EC2 Availability Zones, and you don't need to recover using snapshots or other methods.

- With store files in Amazon S3, you can size your Amazon EMR cluster for your compute requirements instead of data requirements, with 3x replication in HDFS.
- Using Amazon EMR version 5.7.0 or later, you can set up a read-replica cluster, which allows you to maintain read-only copies of data in Amazon S3. You can access the data from the read-replica cluster to perform read operations simultaneously, and in the event that the primary cluster becomes unavailable.
- In Amazon EMR version 6.2.0 and later, persistent HFile Tracking uses a HBase system table called `hbase:storefile` to directly track the HFile paths used for read operations. This feature is enabled by default and does not require manual migration to be performed.

The following illustration shows the HBase components relevant to HBase on Amazon S3.



Enabling HBase on Amazon S3

You can enable HBase on Amazon S3 using the Amazon EMR console, the AWS CLI, or the Amazon EMR API. The configuration is an option during cluster creation. When you use the console, you choose the setting using **Advanced options**. When you use the AWS CLI, use the `--configurations` option to provide a JSON configuration object. Properties of the configuration object specify the storage mode and the root directory location in Amazon S3. The Amazon S3 location that you specify should be in the same region as your Amazon EMR cluster. Only one active cluster at a time can use the same HBase root directory in Amazon S3. For console steps and a detailed create-cluster example using the AWS CLI, see Creating a cluster with HBase. An example configuration object is shown in the following JSON snippet.

```
{
  "Classification": "hbase-site",
  "Properties": {
    "hbase.rootdir": "s3://my-bucket/my-hbase-rootdir"}
},
{
  "Classification": "hbase",
  "Properties": {
  "hbase.emr.storageMode":"s3"
  }
}
```

Note

If you use an Amazon S3 bucket as the `rootdir` for HBase, you must add a slash at the end of the Amazon S3 URI. For example, you must use `"hbase.rootdir: s3://my-bucket/"`, instead of `"hbase.rootdir: s3://my-bucket"`, to avoid issues.

Using a read-replica cluster

After you set up a primary cluster using HBase on Amazon S3, you can create and configure a read-replica cluster that provides read-only access to the same data as the primary cluster. This is useful when you need simultaneous access to query data or uninterrupted access if the primary cluster becomes unavailable. The read-replica feature is available with Amazon EMR version 5.7.0 and later.

The primary cluster and the read-replica cluster are set up the same way with one important difference. Both point to the same `hbase.rootdir` location. However, the `hbase` classification for the read-replica cluster includes the `"hbase.emr.readreplica.enabled":"true"` property.

For example, given the JSON classification for the primary cluster as shown earlier in the topic, the configuration for a read-replica cluster is as follows:

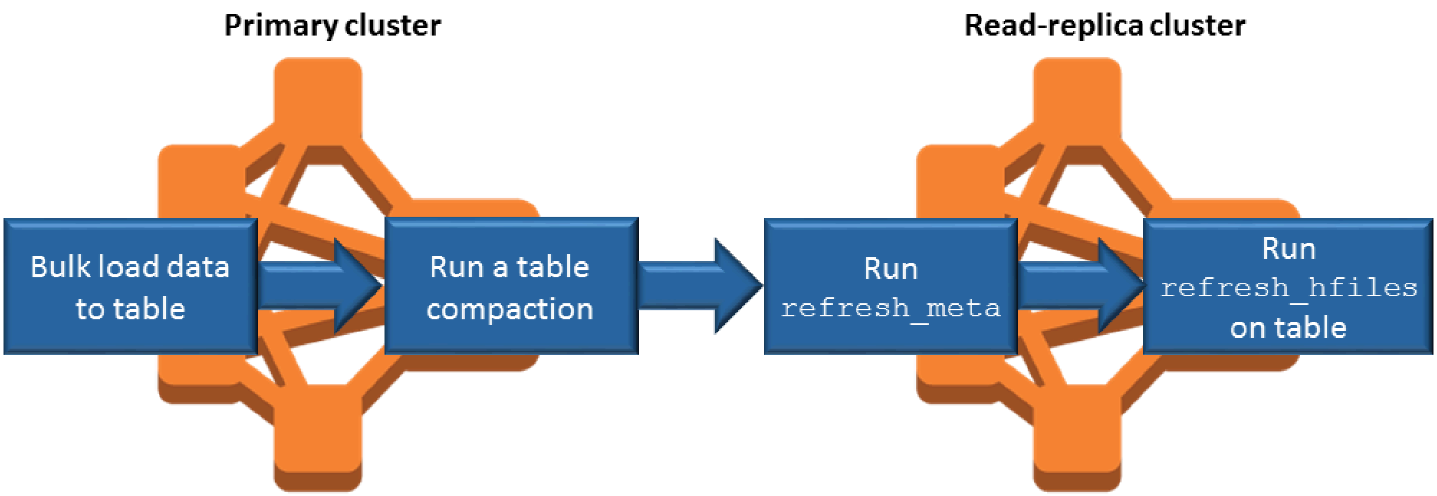
```
{
  "Classification": "hbase-site",
  "Properties": {
    "hbase.rootdir": "s3://my-bucket/my-hbase-rootdir"}
},
{
  "Classification": "hbase",
  "Properties": {
  "hbase.emr.storageMode":"s3",
  "hbase.emr.readreplica.enabled":"true"
  }
}
```

Synchronizing the read replica when you add data

Because the read-replica uses HBase StoreFiles and metadata that the primary cluster writes to Amazon S3, the read-replica is only as current as the Amazon S3 data store. The following guidance can help minimize the lag time between the primary cluster and the read-replica when you write data.

- Bulk load data on the primary cluster whenever possible. For more information, see Bulk loading in Apache HBase documentation.
- A flush that writes store files to Amazon S3 should occur as soon as possible after data is added. Either flush manually or tune flush settings to minimize lag time.
- If compactions might run automatically, run a manual compaction to avoid inconsistencies when compactions are triggered.
- On the read-replica cluster, when any metadata has changed - for example, when HBase region split or compactions occur, or when tables are added or removed - run the `refresh_meta` command.

- On the read-replica cluster, run the `refresh_hfiles` command when records are added to or changed in a table.



Persistent HFile tracking

Persistent HFile tracking uses a HBase system table called `hbase:storefile` to directly track the HFile paths used for read operations. New HFile paths are added to the table as additional data is added to HBase. This removes rename operations as a commit mechanism in the critical write path HBase operations and improves recovery time when opening a HBase region by reading from the `hbase:storefile` system table instead of filesystem directory listing. This feature is enabled by default on Amazon EMR version 6.2.0 and later, and does not require any manual migration steps.

> **Note**
>
> Persistent HFile tracking using the HBase storefile system table does not support the HBase region replication feature. For more information about HBase region replication, see [Timeline-consistent high available reads](#).

Disabling Persistent HFile Tracking

Persistent HFile tracking is enabled by default starting with Amazon EMR release 6.2.0. To disable persistent HFile tracking, specify the following configuration override when launching a cluster:

```
{
  "Classification": "hbase-site",
```

```
  "Properties": {
    "hbase.storefile.tracking.persist.enabled":"false",

 "hbase.hstore.engine.class":"org.apache.hadoop.hbase.regionserver.DefaultStoreEngine"
  }
}
```

> **Note**
>
> When reconfiguring the Amazon EMR cluster, all instance groups must be updated.

Manually Syncing the Storefile Table

The storefile table is kept up to date as new HFiles are created. However, if the storefile table becomes out of sync with the data files for any reason, the following commands can be used to manually sync the data:

Sync storefile table in an online region:

`hbase org.apache.hadoop.hbase.client.example.RefreshHFilesClient <table>`

Sync storefile table in an offline region:

- Remove the storefile table znode.

```
echo "ls /hbase/storefile/loaded" | sudo -u hbase hbase zkcli
[<tableName>, hbase:namespace]
# The TableName exists in the list
echo "delete /hbase/storefile/loaded/<tableName>" | sudo -u hbase hbase zkcli
# Delete the Table ZNode
echo "ls /hbase/storefile/loaded" | sudo -u hbase hbase zkcli
[hbase:namespace]
```

- Assign the region (run in 'hbase shell').

```
hbase cli> assign '<region name>'
```

- If the assignment fails.

```
hbase cli> disable '<table name>'
hbase cli> enable '<table name>'
```

Scaling the Storefile Table

The storefile table is split into four regions by default. If the storefile table is still under heavy write load, the table can be manually split further.

To split a specific hot region, use the following command (run in 'hbase shell').

```
hbase cli> split '<region name>'
```

To split the table, use the following command (run in 'hbase shell').

```
hbase cli> split 'hbase:storefile'
```

Operational considerations

HBase region servers use BlockCache to store data reads in memory and BucketCache to store data reads on local disk. In addition, region servers use MemStore to store data writes in-memory, and use write-ahead logs to store data writes in HDFS before the data is written to HBase StoreFiles in Amazon S3. The read performance of your cluster relates to how often a record can be retrieved from the in-memory or on-disk caches. A cache miss results in the record being read from the StoreFile in Amazon S3, which has significantly higher latency and higher standard deviation than reading from HDFS. In addition, the maximum request rates for Amazon S3 are lower than what can be achieved from the local cache, so caching data may be important for read-heavy workloads. For more information about Amazon S3 performance, see Performance optimization in the *Amazon Simple Storage Service User Guide*.

To improve performance, we recommend that you cache as much of your dataset as possible in EC2 instance storage. Because the BucketCache uses the region server's EC2 instance storage, you can choose an EC2 instance type with a sufficient instance store and add Amazon EBS storage to accommodate the required cache size. You can also increase the BucketCache size on attached instance stores and EBS volumes using the `hbase.bucketcache.size` property. The default setting is 8,192 MB.

For writes, the frequency of MemStore flushes and the number of StoreFiles present during minor and major compactions can contribute significantly to an increase in region server response times. For optimal performance, consider increasing the size of the MemStore flush and HRegion block multiplier, which increases the elapsed time between major compactions, but also increases the lag in consistency if you use a read-replica. In some cases, you may get better performance using larger file block sizes (but less than 5 GB) to trigger Amazon S3 multipart upload functionality in

EMRFS. Amazon EMR's block size default 128 MB. For more information, see HDFS configuration. We rarely see customers who exceed 1 GB block size while benchmarking performance with flushes and compactions. Additionally, HBase compactions and region servers perform optimally when fewer StoreFiles need to be compacted.

Tables can take a significant amount of time to drop on Amazon S3 because large directories need to be renamed. Consider disabling tables instead of dropping.

There is an HBase cleaner process that cleans up old WAL files and store files. With Amazon EMR release version 5.17.0 and later, the cleaner is enabled globally, and the following configuration properties can be used to control cleaner behavior.

Configuration property	Default value	Description
`hbase.regionserver.hfilecleaner.large.thread.count`	1	The number of threads allocated to clean expired large HFiles.
`hbase.regionserver.hfilecleaner.small.thread.count`	1	The number of threads allocated to clean expired small HFiles.
`hbase.cleaner.scan.dir.concurrent.size`	Set to one quarter of all available cores.	The number of threads to scan the oldWALs directories.
`hbase.oldwals.cleaner.thread.size`	2	The number of threads to clean the WALs under the oldWALs directory.

With Amazon EMR 5.17.0 and earlier, the cleaner operation can affect query performance when running heavy workloads, so we recommend that you enable the cleaner only during off-peak times. The cleaner has the following HBase shell commands:

- `cleaner_chore_enabled` queries whether the cleaner is enabled.

- `cleaner_chore_run` manually runs the cleaner to remove files.
- `cleaner_chore_switch` enables or disables the cleaner and returns the previous state of the cleaner. For example, `cleaner_chore_switch true` enables the cleaner.

Properties for HBase on Amazon S3 performance tuning

The following parameters can be adjusted to tune the performance of your workload when you use HBase on Amazon S3.

Configuration property	Default value	Description
`hbase.bucketcache.size`	8,192	The amount of disk space, in MB, reserved on region server Amazon EC2 instance stores and EBS volumes for BucketCache storage. The setting applies to all region server instances. Larger BucketCache sizes generally correspond to improved performance
`hbase.hregion.memstore.flush.size`	134217728	The data limit, in bytes, at which a memstore flush to Amazon S3 is triggered.
`hbase.hregion.memstore.block.multiplier`	4	A multiplier that determines the MemStore upper limit at which updates are blocked. If the MemStore exceeds `hbase.hregion.memstore.flush.size` multiplied by this value, updates are blocked. MemStore flushes and compaction may happen to unblock updates.

Configuration property	Default value	Description
`hbase.hstore.blockingStoreFiles`	10	The maximum number of StoreFiles that can exist in a store before updates are blocked.
`hbase.hregion.max.filesize`	10737418240	The maximum size of a region before the region is split.

Shutting down and restoring a cluster without data loss

To shut down an Amazon EMR cluster without losing data that hasn't been written to Amazon S3, you should flush your MemStore cache to Amazon S3 to write new store files. First, you'll need to disable all tables. The following step configuration can be used when you add a step to the cluster. For more information, see Work with steps using the AWS CLI and console in the *Amazon EMR Management Guide*.

```
Name="Disable all tables",Jar="command-runner.jar",Args=["/bin/bash","/usr/lib/hbase/
bin/disable_all_tables.sh"]
```

Alternatively, you can run the following bash command directly.

```
bash /usr/lib/hbase/bin/disable_all_tables.sh
```

After disabling all tables, flush the `hbase:meta` table using the HBase shell and the following command.

```
flush 'hbase:meta'
```

Then, you can run a shell script provided on the Amazon EMR cluster to flush the MemStore cache. You can either add it as a step or run it directly using the on-cluster AWS CLI. The script disables all HBase tables, which causes the MemStore in each region server to flush to Amazon S3. If the script completes successfully, the data persists in Amazon S3 and the cluster can be terminated.

To restart a cluster with the same HBase data, specify the same Amazon S3 location as the previous cluster either in the AWS Management Console or using the `hbase.rootdir` configuration property.

Write-ahead logs (WAL) for Amazon EMR

With Amazon EMR 6.15 and higher, you can write your Apache HBase write-ahead logs (WAL) to the Amazon EMR WAL. With lower Amazon EMR releases, when you create a cluster with the **HBase on Amazon S3** option, WAL is the only Apache HBase component that gets stored in the local disk for clusters, and you can store other components such as the root directory, store files (HFiles), table metadata, and data on Amazon S3.

You can use Amazon EMR WAL to recover data that didn't flush to Amazon S3. To fully back up your HBase clusters, opt in to use the Amazon EMR WAL service. Behind the scenes, `RegionServer` writes your HBase write-ahead logs (WAL) to the WAL for Amazon EMR.

In the event that your cluster or the AZ becomes unhealthy or unavailable, you can create a new cluster, point it to the same S3 root directory and Amazon EMR WAL workspace, and automatically recover the data in WAL within a few minutes. For more information, see Restoring from Amazon EMR WAL.

> **Note**
>
> Amazon EMR retains your write-ahead log and its data for 30 days from the time you create your cluster. After 30 days, Amazon EMR automatically deletes your Amazon EMR WAL and its data. However, if you launch a new WAL-enabled cluster from the same S3 root directory, you can extend the use of your WAL for 30 days from the launch time of the new cluster. Amazon EMR will still clean up any WAL data from the first cluster after the initial 30-day period. For more information, see Restoring from Amazon EMR WAL.

The following sections describe how to set up and use Amazon EMR WAL with your HBase-enabled EMR cluster.

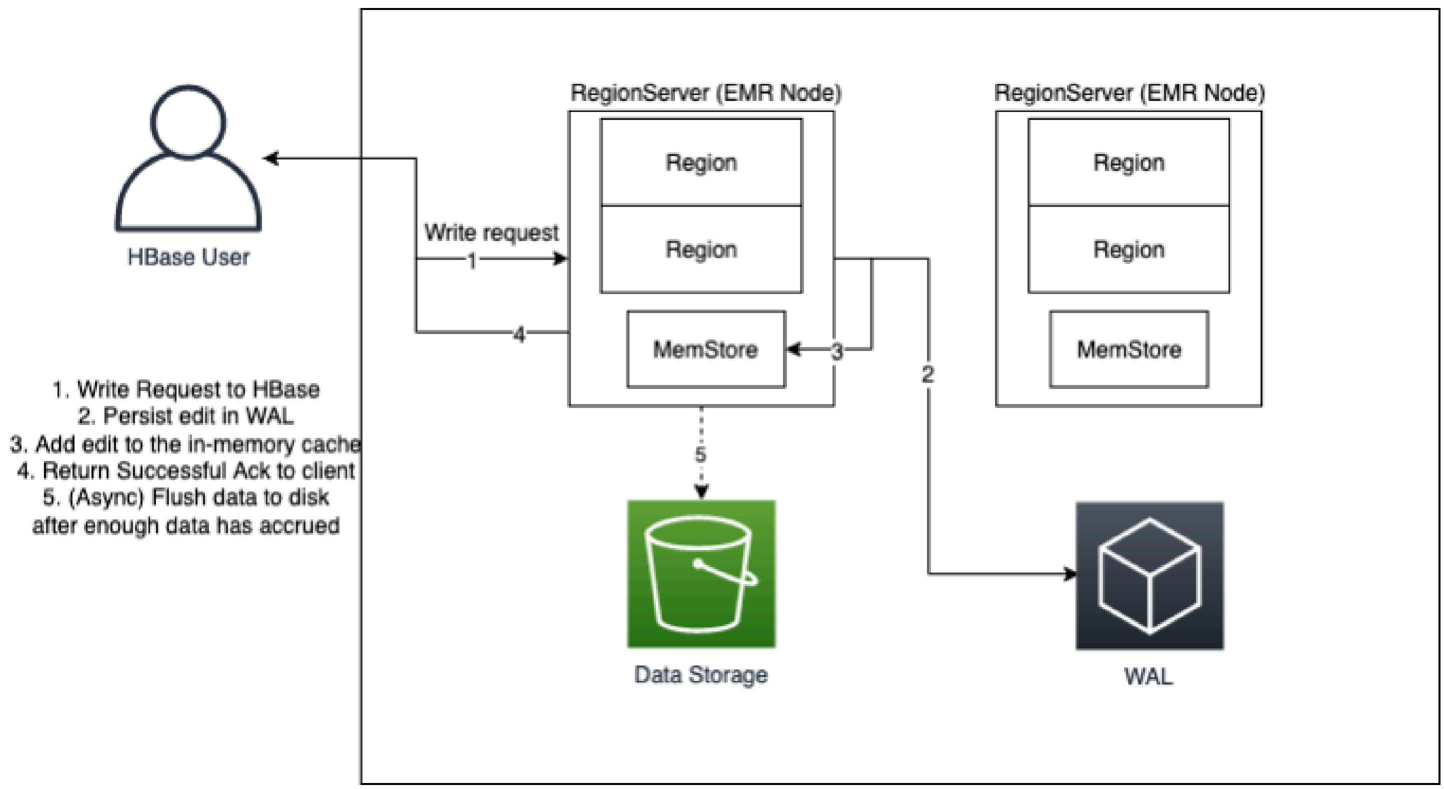
Topics

- Access Amazon EMR WAL through AWS PrivateLink
- Understanding Amazon EMR WAL pricing and metrics
- Tagging WAL workspaces
- Considerations and Regions for Amazon EMR WAL
- Amazon EMR WAL (EMRWAL) CLI reference

Amazon EMR WAL workspaces

Amazon EMR WAL adds the concept of WAL workspaces. A *WAL workspace* is a logical container of WALs. Each write-ahead log in Amazon EMR WAL is encapsulated by a WAL workspace. An EMR cluster writes WALs to exactly one WAL workspace that you configure at cluster launch, or to the `defaultWALworkspace` if you don't specify a workspace. WAL workspaces are not related to any existing HBase terminologies such as namespaces.

You can use WAL workspaces to scope down Amazon EMR WAL IAM permissions to only include the workspaces that the cluster needs to access. You can also tag your WAL workspace for tag-based access control. For more information on tagging, see Tagging WAL workspaces.



Required permissions for Amazon EMR WAL

For your cluster to connect to Amazon EMR WAL, the instance profile for the cluster requires certain IAM permissions:

- Amazon EMR WAL uses the AWSServiceRoleForEMRWAL service-linked role to retrieve a cluster status. Amazon EMR automatically creates this service-linked role when you create a WAL workspace, or HBase will create the service-linked role when you configure a workspace for Amazon EMR WAL and the service-linked role doesn't yet exist.

  Before you can enable Amazon EMR WAL for a cluster, you must configure the permissions to allow automatic creation of the AWSServiceRoleForEMRWAL service-linked role. For more information and an example statement that adds this capability, see Using service-linked roles for write-ahead logging.
- Because Amazon EMR WAL uses HBase Write Ahead Log (WAL), your clusters must use HBase WAL. The following are the minimum IAM permissions that you need to run HBase. Add these to the permissions policy for your instance profile:

```
emrwal:DeleteWal
emrwal:CreateWal
emrwal:CreateWorkspace
emrwal:AppendEdit
emrwal:ReplayEdits
emrwal:GetCurrentWalTime
emrwal:CompleteWalFlush
```

> **Note**
>
> If you scope permissions for Amazon EMR WAL to only the minimal set, some EMRWAL CLI commands won't have the necessary permissions to run.

Enabling Amazon EMR WAL

Use the following steps to enable writing to the Amazon EMR WAL when you create a cluster with the AWS Command Line Interface.

Note

You can't enable Amazon EMR WAL for a cluster that is already running, and you can't launch two clusters with the same S3 root directory. For more information, see Considerations and Regions for Amazon EMR WAL.

1. Before you can create an Amazon EMR WAL enabled cluster, you must add the required permissions to the instance profile that you plan to use with your cluster. For more information, see Required permissions for Amazon EMR WAL.
2. Create a cluster from the AWS CLI. Use the `--configurations` option to provide a JSON configuration object that specifies the `hbase.emr.wal.enabled` property, as shown in the example below.
   - Specify the storage mode and the root directory location in Amazon S3. The Amazon S3 location that you specify should be in the same Region as your EMR cluster, but only one active cluster can use the same HBase root directory in S3 at a time.
   - Create your cluster with the instance groups configuration. You can't use Amazon EMR WAL with the instance fleets configuration. For more information on creating clusters with instance groups, see Configure uniform instance groups in the *Amazon EMR Management Guide*.
   - For console steps to create a cluster, and a detailed `create-cluster` example that uses the AWS CLI, see Creating a cluster with HBase.
3. To enable WAL for the new cluster, set the `hbase.emr.wal.enabled` property to `true`. The following command contains a JSON snippet with an example configuration object.

```
aws emr create-cluster --name "hbasewal" --release-label emr-6.x.y \
--applications Name=HBase --use-default-roles --ec2-attributes KeyName=myKey \
--instance-type m6i.xlarge --instance-count 1 --configurations hbase.json
$cat hbase.json
[
    {
        "Classification": "hbase-site",
        "Properties": {
            "hbase.rootdir": "s3://MyBucket/MyHBaseStore"
        }
    },
    {
```

```
        "Classification": "hbase",
        "Properties": {
            "hbase.emr.storageMode": "s3",
            "hbase.emr.wal.enabled": "true"
        }
    }
]
```

When HBase is online on the newly created cluster, HBase will automatically write WAL data to the Amazon EMR WAL and use the Amazon EMR WAL for recovery purposes.

Example 1: Creating an EMR cluster that uses Amazon EMR WAL

```
[
    {
        "Classification": "hbase-site",
        "Properties": {
            "hbase.rootdir": "s3://MyBucket/MyHBaseStore"
        }
    },
    {
        "Classification": "hbase",
        "Properties": {
            "hbase.emr.storageMode": "s3",
            "hbase.emr.wal.enabled": "true"
        }
    }
]
```

Example 2: Creating an EMR cluster with a custom WAL workspace

```
[
    {
        "Classification": "hbase-site",
        "Properties": {
            "hbase.rootdir": "s3://MyBucket/MyHBaseStore",
            "emr.wal.workspace": "customWorkspaceName"
        }
    },
    {
        "Classification": "hbase",
        "Properties": {
```

```
            "hbase.emr.storageMode": "s3",
            "hbase.emr.wal.enabled": "true"
        }
    }
]
```

Restoring from Amazon EMR WAL

Because the Amazon EMR WAL for your original cluster is retained for 30 days, you can restore and reuse the WAL for a newly-created cluster within that 30-day period. When you launch a new cluster from the same S3 root directory, the 30-day clock restarts from the launch time of the new cluster, as long as the prior 30-day period hasn't expired.

Use the following procedure to restore an existing WAL with a new cluster. This process assumes that you created your original cluster with Amazon EMR WAL enabled.

1. Within 30 days of creating a WAL-enabled cluster, create a new cluster in the same AWS Region as the original cluster. The new cluster can be in the same AZ or in a different AZ within the same Region that the original cluster was created.

   Configure the object properties to specify the storage mode and the root directory location in Amazon S3. The Amazon S3 location that you specify should be in the same Region as your EMR cluster, but only one active cluster can use the same HBase root directory in S3 at a time.

   For console steps to create a cluster, and a detailed `create-cluster` example that uses the AWS CLI, see Creating a cluster with HBase.
2. To use the existing Amazon EMR WAL for the new cluster, set the `hbase.emr.wal.enabled` property to `true`. The following JSON snippet shows an example configuration object.

```
[
    {
        "Classification": "hbase-site",
        "Properties": {
            "hbase.rootdir": "s3://MyBucket/MyHBaseStore"
        }
    },
    {
        "Classification": "hbase",
        "Properties": {
            "hbase.emr.storageMode": "s3",
```

```
            "hbase.emr.wal.enabled": "true"
        }
    }
]
```

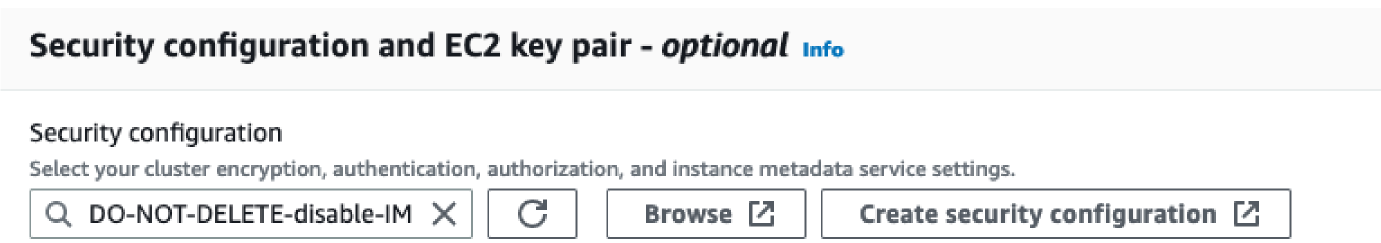
Using security configurations with Amazon EMR WAL

Amazon EMR automatically encrypts both data in transit between your cluster and Amazon EMR WAL service, and the data at rest in Amazon EMR WAL. For more information, see Encryption at rest for Amazon EMR WAL. You can also use a security configuration to bring your own keys from AWS Key Management Service (KMS) service and encrypt the data that you store in Amazon EMR WAL.

Use one of the following methods to select a security configuration when you create a cluster:

Console

From the AWS Management Console, specify the configuration under **Security configuration and EC2 key pair**.



CLI

From the AWS CLI, set the `--security-configuration` parameter when you use the create-cluster command.

For more information, see Encryption at rest for Amazon EMR WAL and Use security configurations to set up cluster security in the *Amazon EMR Management Guide*.

For more security-related information about WAL, see Using service-linked roles for write-ahead logging.

Access Amazon EMR WAL through AWS PrivateLink

If you want to keep your connection within the AWS network, Amazon EMR WAL offers AWS PrivateLink support. To set up AWS PrivateLink, use the AWS Management Console or AWS

Command Line Interface (AWS CLI) to create an interface VPC endpoint that connects to Amazon EMR WAL. For more information, see [Access an AWS service using an interface VPC endpoint](#) in the *AWS PrivateLink Guide*.

The basic steps are as follows:

1. Use the Amazon VPC Console to [create a VPC endpoint](#). Select **Endpoints** and then **Create endpoint**.
2. Keep the Service category as **AWS services**.
3. In the search bar for the **Services** panel, type **`emrwal`**, and then select the service labeled `com.amazonaws.region.emrwal.prod`.
4. Select your VPC and save the endpoint. Make sure that you attach the same security groups to the VPC endpoint that you attach to the EMR cluster.
5. If you want to, you can now enable private DNS hostnames for your new endpoint. Set **Enable DNS hostnames** and **Enable DNS Support** to `true` for your VPC. Then, select your endpoint ID, choose **Edit VPC settings** from the **Actions** menu, and enable private DNS names.
   - The private DNS hostnames for the endpoint will follow the format `prod.emrwal.region.amazonaws.com`.
   - If you don't enable private DNS hostnames, Amazon VPC provides a DNS endpoint name for you in the format `endpointID.prod.emrwal.region.vpce.amazonaws.com`.
6. To use your AWS PrivateLink endpoint, modify the `emr.wal.client.endpoint` configuration when you create your [Amazon EMR WAL enabled cluster](#) as shown in the following example:

```
[
    {
        "Classification": "hbase-site",
        "Properties": {
            "hbase.rootdir": "s3://MyBucket/MyHBaseStore",
            "emr.wal.workspace": "customWorkspaceName",
            "emr.wal.client.endpoint": "https://prod.emrwal.region.amazonaws.com"
        }
    },
    {
        "Classification": "hbase",
        "Properties": {
            "hbase.emr.storageMode": "s3",
            "hbase.emr.wal.enabled": "true"
```

```
        }
    }
]
```

You can also use VPCE policy to allow or restrict access to the Amazon EMR WAL APIs. For more information, see [Control access to VPC endpoints using endpoint policies](#) in the *AWS PrivateLink Guide*.

Understanding Amazon EMR WAL pricing and metrics

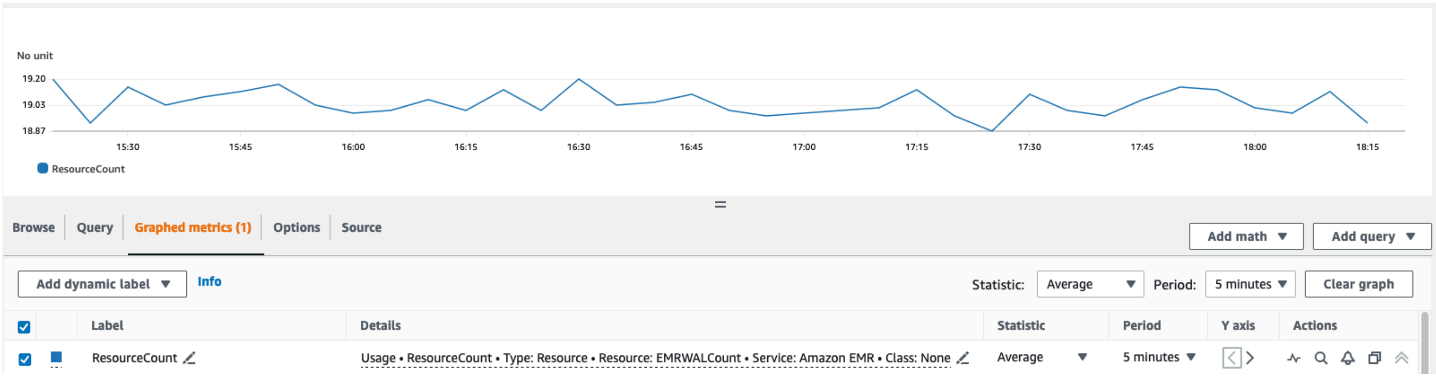
Core feature billing unit	Details
EMR-WAL-Read-GiB	API calls to read data from your table are billed as ReadRequestGiB. This includes `Get and Scan` operations. Reads are charged based on the sizes of the read items. Amazon EMR bills at a minimum of 1 byte. For example, if you read a 1234.12 bytes item, you're charged for 1235 bytes. Reads are aggregated every hour for billing and shown as GiBs.
EMR-WAL-Write-GiB	API calls to write data from your table are billed as Write-GiB. This includes `Put` operations. Writes are charged based on the sizes of the written items. Amazon EMR bills at a minimum of 1 byte. For example, if you write a 1234.12 bytes item, you're charged for 1235 bytes. Writes are aggregated every hour for billing and shown as GiBs.

EMR-WAL-WALHours	The number of WALs that you store on the service are billed as `EMR-WAL-WALHours` . Amazon EMR creates one WAL per HBase Region. For example, if you create 20 HBase tables including system tables, and each table has two HBase Regions, then you use 28,800 WAL hours, calculated as:
	`  20 tables`
	`x  2 Regions per table`
	`x  1 WAL per Region`

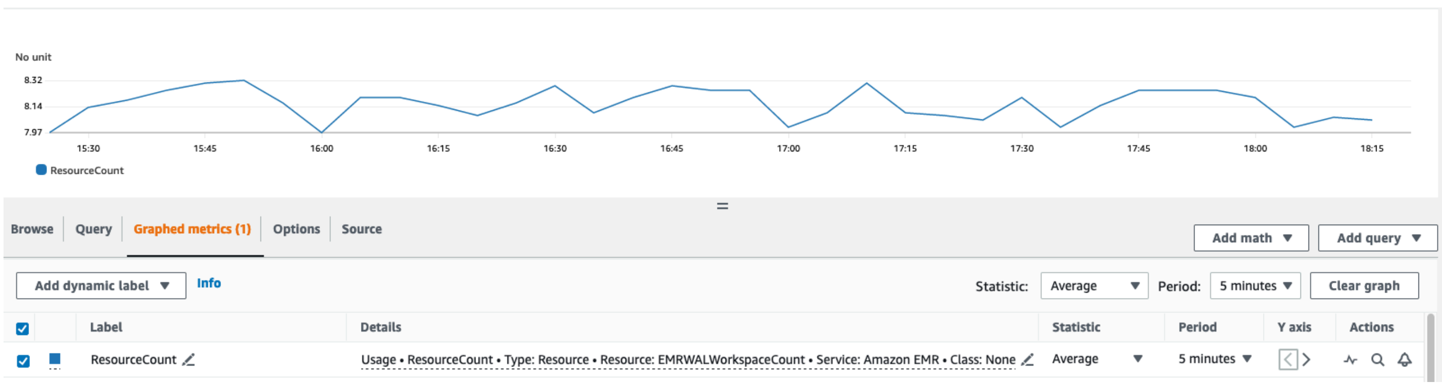
Core feature billing unit	Details

	`x 30 days`
	`x 24 hours`
	`-----------`
	`28,800 EMR-WAL-WALHours`

Example EMRWALCount:



Example EMRWALWorkspaceCount:



Tagging WAL workspaces

You can add tags to a workspace when you create a new workspace and you can add, remove, or list tags from an active workspace for a running cluster. You can't tag the individual resources in the workspace, and you can't update existing tags; instead, remove unwanted tags from the workspace and replace them.

You can tag workspaces from the EMRWAL CLI. For a list of EMRWAL CLI commands for tagging workspaces, see Amazon EMR WAL (EMRWAL) CLI reference.

The following example IAM policy illustrates a scenario that allows workspace CRUDL operations only with the proper tagging key `resource_tag_allow_test_key` and value `resource_tag_allow_test_value`:

```
{
  "Version": "2012-10-17",
  "Statement": [
      {
          "Action": [
              "EMRWAL:*"
          ],
          "Effect": "Allow",
          "Resource": [
              "*",
              "*"
          ],
          "Condition": {
              "StringEquals": {
                  "aws:ResourceTag/resource_tag_allow_test_key": [
                      "resource_tag_allow_test_value"
                  ]
              }
          }
      }
  ]
}
```

To verify that the tag is now required for workspace operations, use the Amazon EMR WAL (EMRWAL) CLI reference to call the `listTagsForResource` command on `tagAllowResourceTag` for the workspace with the desired resource tag. If you configured the condition correctly, the command will succeed.

```
emrwal listTagsForResource -r us-east-1 -arn arn:aws:emrwal:us-east-1:arn:workspace/
tagAllowResourceTag
Tag(Key=resource_tag_allow_test_key, Value=resource_tag_allow_test_value)
```

Considerations and Regions for Amazon EMR WAL

Considerations for Amazon EMR WAL

The following list describes important considerations and limitations of Amazon EMR WAL:

- Amazon EMR WAL is available to use with Amazon EMR releases 6.15.0 and higher.
- Amazon EMR WAL is an opt-in, paid service. You pay for what you use: reads, writes, and data storage. For more information, see Understanding Amazon EMR WAL pricing and metrics and the Amazon EMR pricing page.
- Amazon EMR WAL uses HBase Write Ahead Log (WAL). To use Amazon EMR WAL, your clusters must use HBase WAL.
- To enable Amazon EMR WAL when you create a cluster, you must have the required role permissions. For more information, see Using service-linked roles for write-ahead logging.
- You must enable Amazon EMR WAL when you create the cluster with the AWS Management Console, AWS CLI, or API, and you must use the *instance groups* configuration. You can't enable Amazon EMR WAL in a running cluster if you didn't create the cluster with Amazon EMR WAL. You also can't edit the `hbase-site` configurations to enable Amazon EMR WAL in a running cluster.
- You can only enable Amazon EMR WAL on clusters that use Amazon S3 for the root directory.
- You can't have multiple active clusters on the same HBase root directory in Amazon S3.
- You can't enable Amazon EMR WAL on read replica clusters.
- WAL is replicated across Availability Zones inside the managed service.
- WAL outlives the the cluster, and remains available for the next cluster.
- You can't disable Amazon EMR WAL during launch or when your cluster is operational (in a running state).
- For information on WAL and workspace limits, see Amazon EMR endpoints and quotas.

Region availability for Amazon EMR WAL

Amazon EMR WAL service is available in the following AWS Regions:

- `ap-northeast-1` – Asia Pacific (Tokyo)
- `ap-southeast-1` – Asia Pacific (Singapore)
- `ap-south-1` – Asia Pacific (Mumbai)
- `ap-southeast-2` – Asia Pacific (Sydney)
- `eu-central-1` – Europe (Frankfurt)
- `eu-north-1` – Europe (Stockholm)
- `eu-west-1` – Europe (Ireland)

- `us-east-1` – US East (N. Virginia)
- `us-east-2` – US East (Ohio)
- `us-west-2` – US West (Oregon)
- `ap-northeast-2` – Asia Pacific (Seoul)

Amazon EMR WAL (EMRWAL) CLI reference

The *EMRWAL Command Line Interface (EMRWAL CLI)* is a unified tool to manage your write-ahead log (WAL) for Amazon EMR. The EMRWAL CLI ships with EMR clusters when you enable WAL at the time that you create a cluster. For more information about enabling WAL, see Write-ahead logs (WAL) for Amazon EMR.

The EMRWAL CLI includes the following commands:

Topics

`createWorkspace`

The `createWorkspace` command creates a new Amazon EMR WAL workspace.

Usage:

```
emrwal createWorkspace [-tags <tags>] [-e {endpoint}] [-r {Region}] -w {workspacename}
 [-h]
```

Example:

```
emrwal createWorkspace -w examplews
```

deleteWal

The deleteWals command deletes the Amazon EMR WAL that you specify.

Usage:

```
emrwal deleteWal [-e {endpoint}] [-r {Region}] [-w {workspacename}] [-p <tablePrefix>]
 [-n <walName>] [-N <fullName>] [-R] [-m] [-h]
```

Example:

```
emrwal deleteWal -w examplews -p hbasetable -n examplewal
```

deleteWorkspace

The deleteWorkspace command deletes the Amazon EMR WAL workspace that you specify.

Usage:

```
emrwal deleteWorkspace [-e {endpoint}] [-r {Region}] -w {workspacename} [-h]
```

Example:

```
emrwal deleteWorkspace -w examplews
```

listTagsForResource

The listTagsForResource command lists all of the key-value pair tags for the Amazon EMR WAL workspace that you specify.

Usage:

```
emrwal listTagsForResource -arn {resource-arn} [-e {endpoint}] [-r {Region}] [-h]
```

Example:

```
emrwal listTagsForResource -arn arn:aws:emrwal::1234567891234:workspace/examplews
```

listWals

The `listWals` command lists all of the Amazon EMR WALs in the workspace that you specify.

Usage:

```
emrwal listWals [-nextToken {token-string}] [-pageSize {integer}] [-e {endpoint}] [-
r {Region}] [-w {workspacename}] [-p <tablePrefix>] [-M {integer}] [-h]
```

Example:

```
emrwal listWals -w examplews
```

listWorkspaces

The `listWorkspaces` command lists all of the Amazon EMR WAL workspaces that are available to you.

Usage:

```
emrwal listWorkspaces [-nextToken {token-string}] [-pageSize {integer}] [-e {endpoint}]
 [-r {Region}] [-M {integer}] [-h]
```

Example:

```
emrwal listWorkspaces
```

tagResource

The `tagResource` command assigns one or more key-value pair tags to the Amazon EMR WAL workspace that you specify.

Usage:

```
emrwal tagResource -arn {resource-arn} -tags <tags> [-e {endpoint}] [-r {Region}] [-h]
```

Example:

```
emrwal tagResource -arn arn:aws:emrwal::1234567891234:workspace/examplews -
tags tag_key=tag_value
```

untagResource

The `untagResource` command unassigns one or more key-value pair tags to the Amazon EMR WAL workspace that you specify.

Usage:

```
emrwal untagResource -arn {resource-arn} -tagKeys <tagKeys> [-e {endpoint}] [-
r {Region}] [-h]
```

Example:

```
emrwal untagResource -arn arn:aws:emrwal::1234567891234:workspace/examplews -
tagKeys tag_key
```

Using the HBase shell

After you create an HBase cluster, the next step is to connect to HBase so you can begin reading and writing data (data writes are not supported on a read-replica cluster). You can use the HBase shell to test commands.

To open the HBase shell

1. Use SSH to connect to the main server in the HBase cluster. For information about how to connect to the Amazon EMR primary node using SSH, see Connect to the primary node using SSH in the *Amazon EMR Management Guide*.
2. Run `hbase shell`. The HBase shell opens with a prompt similar to the following.

   ```
   hbase(main):001:0>
   ```

You can issue HBase shell commands from the prompt. For more information about the shell commands and how to call them, type help at the HBase prompt and press Enter.

Create a table

The following command creates a table named 't1' that has a single column family named 'f1'.

```
hbase(main):001:0>create 't1', 'f1'
```

Put a value

The following command puts value 'v1' for row 'r1' in table 't1' and column 'f1'.

```
hbase(main):001:0>put 't1', 'r1', 'f1:col1', 'v1'
```

Get a value

The following command gets the values for row 'r1' in table 't1'.

```
hbase(main):001:0>get 't1', 'r1'
```

Delete a table

The following command drops and deletes table 't1'.

```
hbase(main):001:0>drop 'ns1:t1',false
```

The boolean value corresponds to whether or not you want to archive your table, so you can set it to `true` if you want to save it. You can also run `drop 'ns1:t1'` with no boolean to archive the table.

Access HBase tables with Hive

HBase and Apache Hive are tightly integrated, allowing you run massively parallel processing workloads directly on data stored in HBase. To use Hive with HBase, you can usually launch them on the same cluster. You can, however, launch Hive and HBase on separate clusters. Running HBase and Hive separately on different clusters can improve performance because this allows each application to use cluster resources more efficiently.

The following procedures show how to connect to HBase on a cluster using Hive.

Note

You can only connect a Hive cluster to a single HBase cluster.

To connect Hive to HBase

1. Create separate clusters with Hive and HBase installed or create a single cluster with both HBase and Hive installed.
2. If you are using separate clusters, modify your security groups so that HBase and Hive ports are open between these two primary nodes.
3. Use SSH to connect to the primary node for the cluster with Hive installed. For more information, see Connect to the primary node using SSH in the *Amazon EMR Management Guide*.
4. Launch the Hive shell with the following command.

```
hive
```

5. (Optional) You do not need to do this if HBase and Hive are located on the same cluster. Connect the HBase client on your Hive cluster to the HBase cluster that contains your data. In the following example, *public-DNS-name* is replaced by the public DNS name of the primary node of the HBase cluster, for example: `ec2-50-19-76-67.compute-1.amazonaws.com`.

```
set hbase.zookeeper.quorum=public-DNS-name;
```

6. Proceed to run Hive queries on your HBase data as desired or see the next procedure.

To access HBase data from Hive

- After the connection between the Hive and HBase clusters has been made (as shown in the previous procedure), you can access the data stored on the HBase cluster by creating an external table in Hive.

  The following example, when run from the Hive prompt on the primary node, creates an external table that references data stored in an HBase table called `inputTable`. You can then reference `inputTable` in Hive statements to query and modify data stored in the HBase cluster.

```
set hbase.zookeeper.quorum=ec2-107-21-163-157.compute-1.amazonaws.com;

create external table inputTable (key string, value string)
     stored by 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
      with serdeproperties ("hbase.columns.mapping" = ":key,f1:col1")
      tblproperties ("hbase.table.name" = "t1");

select count(key) from inputTable ;
```

For a more advanced use case and example combining HBase and Hive, see the AWS Big Data Blog post, Combine NoSQL and massively parallel analytics using Apache HBase and Apache Hive on Amazon EMR.

Using HBase snapshots

HBase uses a built-in snapshot functionality to create lightweight backups of tables. In EMR clusters, these backups can be exported to Amazon S3 using EMRFS. You can create a snapshot on the primary node using the HBase shell. This topic shows you how to run these commands interactively with the shell or through a step using `command-runner.jar` with either the AWS CLI or AWS SDK for Java. For more information about other types of HBase backups, see HBase backup in the HBase documentation.

Create a snapshot using a table

```
hbase snapshot create -n snapshotName -t tableName
```

Using command-runner.jar from the AWS CLI:

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps Name="HBase Shell Step",Jar="command-runner.jar",\
Args=[ "hbase", "snapshot", "create","-n","snapshotName","-t","tableName"]
```

AWS SDK for Java

```
HadoopJarStepConfig hbaseSnapshotConf = new HadoopJarStepConfig()
  .withJar("command-runner.jar")
  .withArgs("hbase","snapshot","create","-n","snapshotName","-t","tableName");
```

Note

If your snapshot name is not unique, the create operation fails with a return code of -1 or 255 but you may not see an error message that states what went wrong. To use the same snapshot name, delete it and then re-create it.

Delete a snapshot

```
hbase shell
>> delete_snapshot 'snapshotName'
```

View snapshot info

```
hbase snapshot info -snapshot snapshotName
```

Export a snapshot to Amazon S3

Important

If you do not specify a `-mappers` value when exporting a snapshot, HBase uses an arbitrary calculation to determine the number of mappers. This value can be very large depending on your table size, which negatively affects running jobs during the export. For this reason, we recommend that you specify the `-mappers` parameter, the `-bandwidth` parameter (which specifies the bandwidth consumption in megabytes per second), or both to limit the cluster resources used by the export operation. Alternatively, you can run the export snapshot operation during a period of low usage.

```
hbase snapshot export -snapshot snapshotName \
-copy-to s3://bucketName/folder -mappers 2
```

Using command-runner.jar from the AWS CLI:

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps Name="HBase Shell Step",Jar="command-runner.jar",\
Args=[ "hbase", "snapshot", "export","-snapshot","snapshotName","-copy-
to","s3://bucketName/folder","-mappers","2","-bandwidth","50"]
```

AWS SDK for Java:

```
HadoopJarStepConfig hbaseImportSnapshotConf = new HadoopJarStepConfig()
  .withJar("command-runner.jar")
  .withArgs("hbase","snapshot","export",
      "-snapshot","snapshotName","-copy-to",
      "s3://bucketName/folder",
      "-mappers","2","-bandwidth","50");
```

Import snapshot from Amazon S3

Although this is an import, the HBase option used here is still export.

```
sudo -u hbase hbase snapshot export \
-D hbase.rootdir=s3://bucketName/folder \
-snapshot snapshotName \
-copy-to hdfs://masterPublicDNSName:8020/user/hbase \
-mappers 2
```

Using command-runner.jar from the AWS CLI:

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps Name="HBase Shell Step",Jar="command-runner.jar", \
Args=["sudo","-u","hbase","hbase snapshot export","-snapshot","snapshotName", \
"-D","hbase.rootdir=s3://bucketName/folder", \
"-copy-to","hdfs://masterPublicDNSName:8020/user/hbase","-mappers","2","-chmod","700"]
```

AWS SDK for Java:

```
HadoopJarStepConfig hbaseImportSnapshotConf = new HadoopJarStepConfig()
  .withJar("command-runner.jar")
  .withArgs("sudo","-u","hbase","hbase","snapshot","export", "-D","hbase.rootdir=s3://
path/to/snapshot",
      "-snapshot","snapshotName","-copy-to",
```

```
    "hdfs://masterPublicDNSName:8020/user/hbase",
    "-mappers","2","-chuser","hbase");
```

Restore a table from snapshots within the HBase shell

```
hbase shell
>> disable tableName
>> restore_snapshot snapshotName
>> enable tableName
```

HBase currently does not support all snapshot commands found in the HBase shell. For example, there is no HBase command-line option to restore a snapshot, so you must restore it within a shell. This means that `command-runner.jar` must run a Bash command.

Note

Because the command used here is `echo`, it is possible that your shell command will still fail even if the command run by Amazon EMR returns a `0` exit code. Check the step logs if you choose to run a shell command as a step.

```
echo 'disable tableName; \
restore_snapshot snapshotName; \
enable tableName' | hbase shell
```

Here is the step using the AWS CLI. First, create the following `snapshot.json` file:

```
[
  {
    "Name": "restore",
    "Args": ["bash", "-c", "echo $'disable \"tableName\"; restore_snapshot
 \"snapshotName\"; enable \"tableName\"' | hbase shell"],
    "Jar": "command-runner.jar",
    "ActionOnFailure": "CONTINUE",
    "Type": "CUSTOM_JAR"
  }
]
```

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
```

```
--steps file://./snapshot.json
```

AWS SDK for Java:

```
HadoopJarStepConfig hbaseRestoreSnapshotConf = new HadoopJarStepConfig()
  .withJar("command-runner.jar")
  .withArgs("bash","-c","echo $'disable \"tableName\"; restore_snapshot \"snapshotName
\"; enable \"snapshotName\"' | hbase shell");
```

Configure HBase

Although the default HBase settings should work for most applications, you can modify your HBase configuration settings. To do this, use properties of HBase configuration classifications. For more information, see Configure applications.

The following example creates a cluster with an alternate HBase root directory based on a configuration file, `myConfig.json`, stored in Amazon S3.

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --release-label emr-7.2.0 --applications Name=HBase \
--instance-type m5.xlarge --instance-count 3 --configurations https://s3.amazonaws.com/
mybucket/myfolder/myConfig.json
```

The `myConfig.json` file specifies the `hbase.rootdir` property for the `hbase-site` configuration classification as shown in the following example. Replace *`ip-XXX-XX-XX-XXX.ec2.internal`* with the internal DNS hostname of the cluster's primary node.

```
[
  {
    "Classification":"hbase-site",
    "Properties": {
       "hbase.rootdir": "hdfs://ip-XXX-XX-XX-XXX.ec2.internal:8020/user/
myCustomHBaseDir"
```

```
      }
    }
]
```

Note

With Amazon EMR version 5.21.0 and later, you can override cluster configurations and specify additional configuration classifications for each instance group in a running cluster. You do this by using the Amazon EMR console, the AWS Command Line Interface (AWS CLI), or the AWS SDK. For more information, see [Supplying a Configuration for an Instance Group in a Running Cluster](#).

Changes to memory allocation in YARN

HBase is not running as a YARN application, thus it is necessary to recalculate the memory allocated to YARN and its applications, which results in a reduction in overall memory available to YARN if HBase is installed. You should take this into account when planning to co-locate YARN applications and HBase on the same clusters. The instance types with less than 64 GB of memory have half the memory available to NodeManager, which is then allocated to the HBase RegionServer. For instance types with memory greater than 64 GB, HBase RegionServer memory is capped at 32 GB. As a general rule, YARN setting memory is some multiple of MapReduce reducer task memory.

The tables in [Default values for task configuration settings](#) show changes to YARN settings based on the memory needed for HBase.

HBase port numbers

Some port numbers chosen for HBase are different from the default. The following are interfaces and ports for HBase on Amazon EMR.

HBase ports

Interface	Port	Protocol
HMaster	16000	TCP
HMaster UI	16010	HTTP

Interface	Port	Protocol
RegionServer	16020	TCP
RegionServer Info	16030	HTTP
REST server	8070	HTTP
REST UI	8085	HTTP
Thrift server	9090	TCP
Thrift server UI	9095	HTTP

Important

The `kms-http-port` is 9700 and the `kms-admin-port` is 9701 in Amazon EMR release version 4.6.0 and later.

HBase site settings to optimize

You can set any or all of the HBase site settings to optimize the HBase cluster for your application's workload. We recommend the following settings as a starting point in your investigation.

zookeeper.session.timeout

The default timeout is 40 seconds (40000 ms). If a region server crashes, this is how long it takes the master server to notice the absence of the region server and start recovery. To help the master server recover faster, you can reduce this value to a shorter time period. The following example uses 30 seconds, or 30000 ms:

```
[
  {
    "Classification":"hbase-site",
    "Properties": {
       "zookeeper.session.timeout": "30000"
    }
  }
```

```
]
```

hbase.regionserver.handler.count

This defines the number of threads the region server keeps open to serve requests to tables. The default of 10 is low, in order to prevent users from killing their region servers when using large write buffers with a high number of concurrent clients. The rule of thumb is to keep this number low when the payload per request approaches the MB range (big puts, scans using a large cache) and high when the payload is small (gets, small puts, ICVs, deletes). The following example raises the number of open threads to 30:

```
[
  {
    "Classification":"hbase-site",
    "Properties": {
       "hbase.regionserver.handler.count": "30"
    }
  }
]
```

hbase.hregion.max.filesize

This parameter governs the size, in bytes, of the individual regions. By default, it is set to `1073741824`. If you are writing a lot of data into your HBase cluster, and it's causing frequent splitting, you can increase this size to make individual regions bigger. It reduces splitting but takes more time to load-balance regions from one server to another.

```
[
  {
    "Classification":"hbase-site",
    "Properties": {
       "hbase.hregion.max.filesize": "1073741824"
    }
  }
]
```

hbase.hregion.memstore.flush.size

This parameter governs the maximum size of memstore, in bytes, before it is flushed to disk. By default, it is `134217728`. If your workload consists of short bursts of write operations, you might want to increase this limit so that all writes stay in memory during the burst and get flushed to disk later. This can boost performance during bursts.

```
[
  {
    "Classification":"hbase-site",
    "Properties": {
       "hbase.hregion.memstore.flush.size": "134217728"
    }
  }
]
```

View the HBase user interface

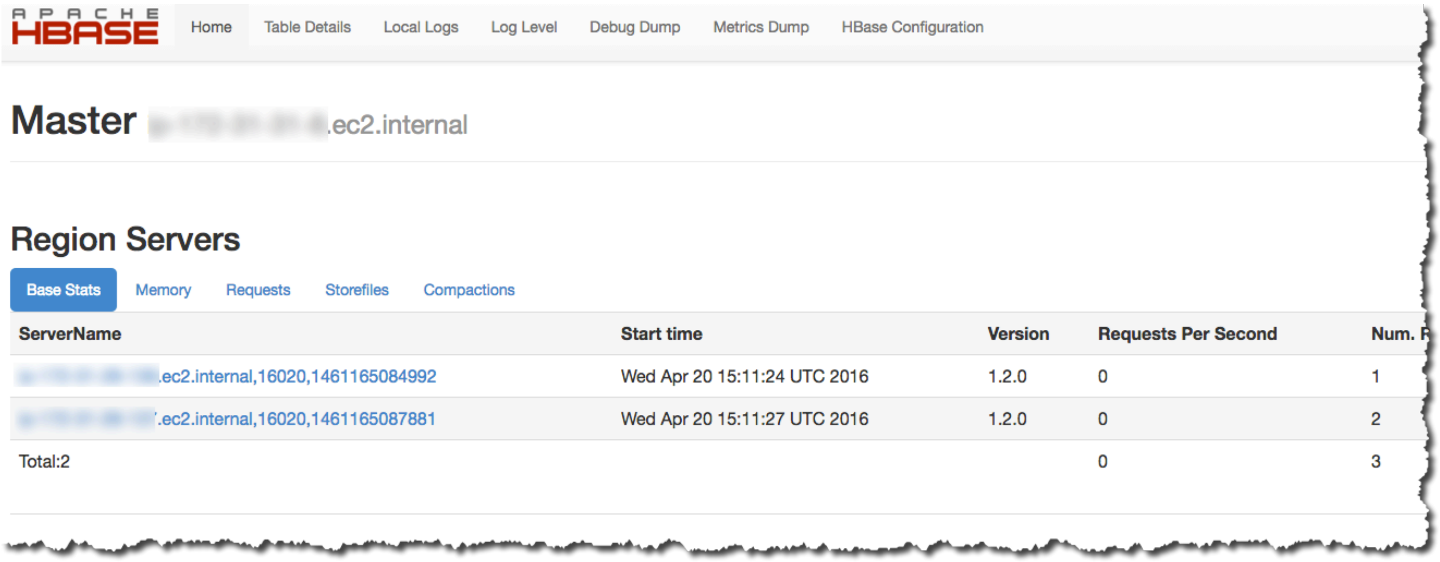
Note

The HBase user interface uses insecure HTTP connections by default. To enable secure HTTP (HTTPS), set the `hbase.ssl.enabled` property for the `hbase-site` classification to `true` in your HBase configuration. For more information about using secure HTTP (HTTPS) for the HBase web UI, see the Apache HBase Reference Guide.

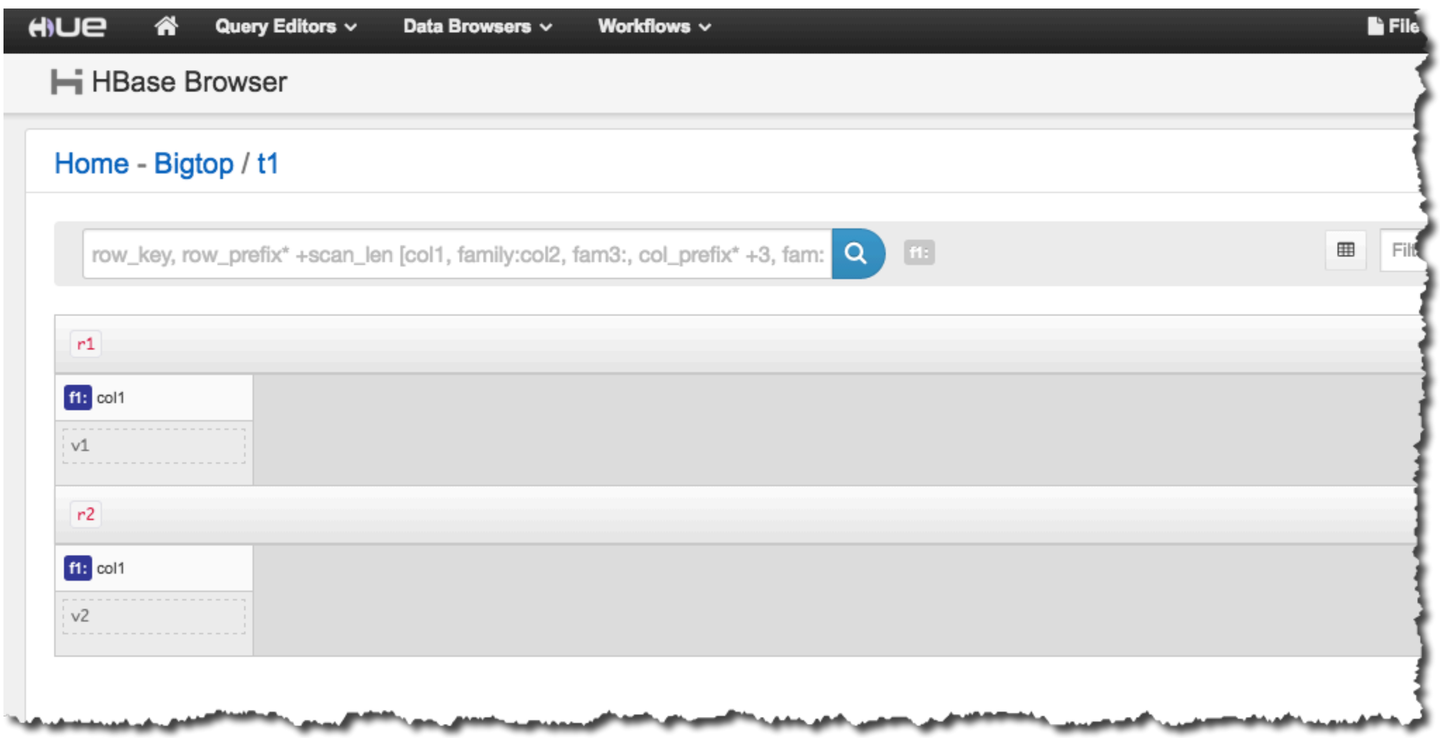
HBase provides a web-based user interface that you can use to monitor your HBase cluster. When you run HBase on Amazon EMR, the web interface runs on the primary node and can be viewed using port forwarding, also known as creating an SSH tunnel.

To view the HBase user interface

1. Use SSH to tunnel into the primary node and create a secure connection. For more information, see Option 2, part 1: Set up an SSH tunnel to the primary node using dynamic port forwarding in the *Amazon EMR Management Guide*.
2. Install a web browser with a proxy tool, such as the FoxyProxy plug-in for Firefox, to create a SOCKS proxy for AWS domains. For more information, see Option 2, part 2: Configure proxy settings to view websites hosted on the primary node in the *Amazon EMR Management Guide*.
3. With the proxy set and the SSH connection open, you can view the HBase UI by opening a browser window with **http://*master-public-dns-name*:16010/master-status**, where *master-public-dns-name* is the public DNS address of the cluster's primary node.



You can also view HBase in Hue. For example, the following shows the table, `t1`, created in [Using the HBase shell](#):



For more information about Hue, see [Hue](#).

View HBase log files

As part of its operation, HBase writes log files with details about configuration settings, daemon actions, and exceptions. These log files can be useful for debugging issues with HBase as well as for tracking performance.

If you configure your cluster to persist log files to Amazon S3, you should know that logs are written to Amazon S3 every five minutes, so there may be a slight delay before the latest log files are available.

To view HBase logs on the primary node

- You can view the current HBase logs by using SSH to connect to the primary node, and navigating to the `/var/log/hbase` directory. These logs are not available after the cluster is terminated unless you enable logging to Amazon S3 when the cluster is launched.

To view HBase logs on Amazon S3

- To access HBase logs and other cluster logs on Amazon S3, and to have them available after the cluster terminates, specify an Amazon S3 bucket to receive these logs when you create the cluster. This is done using the `--log-uri` option. For more information about enabling logging for your cluster, see Configure logging and debugging (optional) in the *Amazon EMR Management Guide*.

Monitor HBase with Ganglia

The Ganglia open-source project is a scalable, distributed system designed to monitor clusters and grids while minimizing the impact on their performance. When you enable Ganglia on your cluster, you can generate reports and view the performance of the cluster as a whole, as well as inspect the performance of individual node instances. For more information about the Ganglia open-source project, see http://ganglia.info/. For more information about using Ganglia with Amazon EMR clusters, see Ganglia.

After the cluster is launched with Ganglia configured, you can access the Ganglia graphs and reports using the graphical interface running on the primary node.

Ganglia stores log files on the primary node in the `/mnt/var/lib/ganglia/rrds/` directory. Earlier release versions of Amazon EMR may store log files in the `/var/log/ganglia/rrds/` directory.

To configure a cluster for Ganglia and HBase using the AWS CLI

- Use a `create-cluster` command similar to the following:

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.2.0 \
--applications Name=HBase Name=Ganglia --use-default-roles \
--ec2-attributes KeyName=myKey --instance-type m5.xlarge \
--instance-count 3
```

> **Note**
>
> If the default Amazon EMR service role and Amazon EC2 instance profile don't exist, an error occurs. Use the `aws emr create-default-roles` command to create them and then try again.

For more information, see Amazon EMR commands in the AWS CLI.

To view HBase metrics in the Ganglia web interface

1. Use SSH to tunnel into the primary node and create a secure connection. For more information, see Option 2, part 1: Set up an SSH tunnel to the primary node using dynamic port forwarding in the *Amazon EMR Management Guide*.
2. Install a web browser with a proxy tool, such as the FoxyProxy plug-in for Firefox, to create a SOCKS proxy for AWS domains. For more information, see Option 2, part 2: Configure proxy settings to view websites hosted on the primary node in the *Amazon EMR Management Guide*.
3. With the proxy set and the SSH connection open, you can view the Ganglia metrics by opening a browser window with http://*master-public-dns-name*/ganglia/, where *master-public-dns-name* is the public DNS address of the master server in the HBase cluster.

To view Ganglia log files on the primary node

- If the cluster is still running, you can access the log files by using SSH to connect to the primary node and navigating to the `/mnt/var/lib/ganglia/rrds/` directory. For EMR 3.x, navigate to the `/var/log/ganglia/rrds` directory. For more information, see Connect to the primary node using SSH in the *Amazon EMR Management Guide*.

To view Ganglia log files on Amazon S3

- The Ganglia log files are not automatically written to Amazon S3 even if you enable logging for your cluster. To view Ganglia log files on Amazon S3, you must manually push the logs from `/mnt/var/lib/ganglia/rrds/` to the S3 bucket.

Migrating from previous HBase versions

To migrate data from a previous HBase version, see Upgrading and HBase version number and compatibility in the Apache HBase Reference Guide. You may need to pay special attention to the requirements for upgrading from pre-1.0 versions of HBase.

HBase release history

The following table lists the version of HBase included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in Amazon EMR 7.x release versions, Amazon EMR 6.x release versions, or Amazon EMR 5.x release versions.

HBase version information

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-7.2.0	2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-

Amazon EMR Release label	HBase Version	Components installed with HBase
		kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server
emr-5.36.2	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-7.1.0	2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-7.0.0	2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-6.15.0	2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-6.14.0	2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-6.13.0	2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-6.12.0	2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-6.11.1	2.4.15	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-6.11.0	2.4.15	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-6.10.1	2.4.15	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-6.10.0	2.4.15	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-6.9.1	2.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-6.9.0	2.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-6.8.1	2.4.12	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-6.8.0	2.4.12	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-6.7.0	2.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.36.1	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.36.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-6.6.0	2.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.35.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.5.0	2.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-6.4.0	2.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.3.1	2.2.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-6.3.0	2.2.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.2.1	2.2.6-amzn-0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-6.2.0	2.2.6-amzn-0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.1.1	2.2.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-6.1.0	2.2.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.0.1	2.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-6.0.0	2.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.34.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.33.1	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.33.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.32.1	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.32.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.31.1	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.31.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.30.2	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.30.1	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.30.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.29.0	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.28.1	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.28.0	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.27.1	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.27.0	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.26.0	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.25.0	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.24.1	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.24.0	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.23.1	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.23.0	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.22.0	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.21.2	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.21.1	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.21.0	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.20.1	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.20.0	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.19.1	1.4.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.19.0	1.4.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.18.1	1.4.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.18.0	1.4.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.17.2	1.4.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.17.1	1.4.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.17.0	1.4.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.16.1	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.16.0	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.15.1	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.15.0	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.14.2	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.14.1	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.14.0	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.13.1	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.13.0	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.12.3	1.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.12.2	1.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.12.1	1.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.12.0	1.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.11.4	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.11.3	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.11.2	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.11.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.11.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.10.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.10.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.9.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.9.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.8.3	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.8.2	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.8.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.8.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.7.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.7.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.6.1	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.6.0	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.5.4	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.5.3	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.5.2	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.5.1	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.5.0	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.4.1	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.4.0	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.3.2	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.3.1	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.3.0	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.2.3	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.2.2	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.2.1	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.2.0	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.1.1	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.1.0	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.0.3	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.0.2	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.0.1	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-5.0.0	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.9.6	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-4.9.5	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.9.4	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-4.9.3	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.9.2	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-4.9.1	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.8.5	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-4.8.4	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.8.3	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-4.8.2	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.8.1	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-4.8.0	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.7.4	1.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-4.7.3	1.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.7.2	1.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-4.7.1	1.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.7.0	1.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	HBase Version	Components installed with HBase
emr-4.6.1	1.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.6.0	1.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Apache HCatalog

HCatalog is a tool that allows you to access Hive metastore tables within Pig, Spark SQL, and/or custom MapReduce applications. HCatalog has a REST interface and command line client that allows you to create tables or do other operations. You then write your applications to access the tables using HCatalog libraries. For more information, see Using HCatalog. HCatalog is included in Amazon EMR release version 4.4.0 and later.

HCatalog on Amazon EMR release version 5.8.0 and later supports using AWS Glue Data Catalog as the metastore for Hive. For more information, see Using AWS Glue Data Catalog as the metastore for Hive.

The following table lists the version of HCatalog included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with HCatalog.

For the version of components installed with HCatalog in this release, see Release 6.15.0 Component Versions.

HCatalog version information for emr-6.15.0

Amazon EMR Release Label	HCatalog Version	Components Installed With HCatalog
emr-6.15.0	HCatalog 3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

The following table lists the version of HCatalog included in the latest release of the Amazon EMR 5.x series, along with the components that Amazon EMR installs with HCatalog.

For the version of components installed with HCatalog in this release, see Release 5.36.2 Component Versions.

HCatalog version information for emr-5.36.2

Amazon EMR Release Label	HCatalog Version	Components Installed With HCatalog
emr-5.36.2	HCatalog 2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Topics

Creating a cluster with HCatalog

Although HCatalog is included in the Hive project, you must install it as its own application.

To launch a cluster with HCatalog installed using the console

The following procedure creates a cluster with HCatalog installed. For more information about creating clusters using the console, including **Advanced Options** see Plan and configure clusters in the *Amazon EMR Management Guide*.

1. Open the Amazon EMR console at https://console.aws.amazon.com/emr.
2. Choose **Create cluster** to use **Quick Create**.
3. For the **Software Configuration** field, choose **Amazon Release Version emr-4.4.0** or later.
4. In the **Select Applications** field, choose either **All Applications** or **HCatalog**.
5. Select other options as necessary and then choose **Create cluster**.

To launch a cluster with HCatalog using the AWS CLI

- Create the cluster with the following command:

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --name "Cluster with Hcat" --release-label emr-7.2.0 \
--applications Name=HCatalog --ec2-attributes KeyName=myKey \
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

Using HCatalog

You can use HCatalog within various applications that use the Hive metastore. The examples in this section show how to create a table and use it in the context of Pig and Spark SQL.

Disable direct write when using HCatalog HStorer

Whenever an application uses HCatStorer to write to an HCatalog table stored in Amazon S3, disable the direct write feature of Amazon EMR. For example, disable direct write when using the Pig `STORE` command or when running Sqoop jobs that write HCatalog tables to Amazon S3. You can disable the direct write feature by setting the `mapred.output.direct.NativeS3FileSystem` and the `mapred.output.direct.EmrFileSystem` configurations to `false`. The following example demonstrates how to set these configurations using Java.

```
Configuration conf = new Configuration();
conf.set("mapred.output.direct.NativeS3FileSystem", "false");
conf.set("mapred.output.direct.EmrFileSystem", "false");
```

Create a table using the HCat CLI and use that data in Pig

Create the following script, `impressions.q`, on your cluster:

```
CREATE EXTERNAL TABLE impressions (
    requestBeginTime string, adId string, impressionId string, referrer string,
    userAgent string, userCookie string, ip string
  )
  PARTITIONED BY (dt string)
  ROW FORMAT
    serde 'org.apache.hive.hcatalog.data.JsonSerDe'
    with serdeproperties ( 'paths'='requestBeginTime, adId, impressionId, referrer,
 userAgent, userCookie, ip' )
  LOCATION 's3://[your region].elasticmapreduce/samples/hive-ads/tables/impressions/';
ALTER TABLE impressions ADD PARTITION (dt='2009-04-13-08-05');
```

Execute the script using the HCat CLI:

```
% hcat -f impressions.q
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-
log4j.properties
OK
Time taken: 4.001 seconds
OK
Time taken: 0.519 seconds
```

Open the Grunt shell and access the data in `impressions`:

```
% pig -useHCatalog -e "A = LOAD 'impressions' USING
 org.apache.hive.hcatalog.pig.HCatLoader();
B = LIMIT A 5;
dump B;"
<snip>
```

```
(1239610346000,m9nwdo67Nx6q2kI25qt5On7peICfUM,omkxkaRpNhGPDucAiBErSh1cs0MThC,cartoonnetwork.com
 (compatible; MSIE 7.0; Windows NT 6.0; FunWebProducts; GTB6; SLCC1; .NET CLR
 2.0.50727; Media Center PC
 5.0; .NET,wcVWWTascoPbGt6bdqDbuWTPPHgOPs,69.191.224.234,2009-04-13-08-05)
(1239611000000,NjriQjdODgWBKnkGJUP6GNTbDeK4An,AWtXPkfaWGOaNeL9OOsFU8Hcj6eLHt,cartoonnetwork.com
 (compatible; MSIE 7.0; Windows NT 5.1; GTB6; .NET CLR
 1.1.4322),OaMU1F2gE4CtADVHAbKjjRRks5kIgg,57.34.133.110,2009-04-13-08-05)
(1239610462000,Irpv3oiu0I5QNQiwSSTIshrLdo9cM1,i1LDq44LRSJF0hbmhB8Gk7k9gMWtBq,cartoonnetwork.com
 (compatible; MSIE 6.0; Windows NT 5.2; SV1; .NET CLR 1.1.4322;
 InfoPath.1),QSb3wkLR4JAIut4Uq6FNFQIR1rCVwU,42.174.193.253,2009-04-13-08-05)
(1239611007000,q2Awfnpe0JAvhInaIp0VGx9KTs0oPO,s3HvTflPB8JIE0IuM6hOEebWWpOtJV,cartoonnetwork.com
 (compatible; MSIE 6.0; Windows NT 5.2; SV1; .NET CLR 1.1.4322;
 InfoPath.1),QSb3wkLR4JAIut4Uq6FNFQIR1rCVwU,42.174.193.253,2009-04-13-08-05)
(1239610398000,c362vpAB0soPKGHRS43cj6TRwNeOGn,jeas5nXbQInGAgFB8jlkhnprN6cMw7,cartoonnetwork.com
 (compatible; MSIE 8.0; Windows NT 5.1; Trident/4.0; GTB6; .NET CLR
 1.1.4322),k96n5PnUmwHKfiUI0TFP0TNMfADgh9,51.131.29.87,2009-04-13-08-05)
7120 [main] INFO  org.apache.pig.Main  - Pig script completed in 7 seconds and 199
 milliseconds (7199 ms)
16/03/08 23:17:10 INFO pig.Main: Pig script completed in 7 seconds and 199 milliseconds
 (7199 ms)
```

Accessing the table using Spark SQL

This example creates a Spark DataFrame from the table created in the first example and shows the first 20 lines:

```
% spark-shell --jars /usr/lib/hive-hcatalog/share/hcatalog/hive-hcatalog-core-1.0.0-
amzn-3.jar
<snip>
scala> val hiveContext = new org.apache.spark.sql.hive.HiveContext(sc);
scala> val df = hiveContext.sql("SELECT * FROM impressions")
scala> df.show()
<snip>
16/03/09 17:18:46 INFO DAGScheduler: ResultStage 0 (show at <console>:32) finished in
 10.702 s
16/03/09 17:18:46 INFO DAGScheduler: Job 0 finished: show at <console>:32, took
 10.839905 s
+----------------+--------------------+--------------------+------------------
+--------------------+--------------------+--------------+-----------------+
|requestbegintime|                adid|        impressionid|          referrer|
   useragent|          usercookie|            ip|                dt|
+----------------+--------------------+--------------------+------------------
+--------------------+--------------------+--------------+-----------------+
```

```
|   1239610346000|m9nwdo67Nx6q2kI25...|omkxkaRpNhGPDucAi...|cartoonnetwork.com|
Mozilla/4.0 (comp...|wcVWWTascoPbGt6bd...|69.191.224.234|2009-04-13-08-05|
|   1239611000000|NjriQjdODgWBKnkGJ...|AWtXPkfaWGOaNeL9O...|cartoonnetwork.com|
Mozilla/4.0 (comp...|OaMU1F2gE4CtADVHA...| 57.34.133.110|2009-04-13-08-05|
|   1239610462000|Irpv3oiu0I5QNQiwS...|i1LDq44LRSJF0hbmh...|cartoonnetwork.com|
Mozilla/4.0 (comp...|QSb3wkLR4JAIut4Uq...|42.174.193.253|2009-04-13-08-05|
|   1239611007000|q2Awfnpe0JAvhInaI...|s3HvTflPB8JIE0IuM...|cartoonnetwork.com|
Mozilla/4.0 (comp...|QSb3wkLR4JAIut4Uq...|42.174.193.253|2009-04-13-08-05|
|   1239610398000|c362vpAB0soPKGHRS...|jeas5nXbQInGAgFB8...|cartoonnetwork.com|
Mozilla/4.0 (comp...|k96n5PnUmwHKfiUI0...|  51.131.29.87|2009-04-13-08-05|
|   1239610600000|cjBTpruoaiEtqLuMX...|XwlohBSs8Ipxs1bRa...|cartoonnetwork.com|
Mozilla/4.0 (comp...|k96n5PnUmwHKfiUI0...|  51.131.29.87|2009-04-13-08-05|
|   1239610804000|Ms3eJHNAEItpxvimd...|4SIj4pGmgVLl625BD...|cartoonnetwork.com|
Mozilla/4.0 (comp...|k96n5PnUmwHKfiUI0...|  51.131.29.87|2009-04-13-08-05|
|   1239610872000|h5bccHX6wJReDi1jL...|EFAWIiBdVfnxwAMWP...|cartoonnetwork.com|
Mozilla/4.0 (comp...|k96n5PnUmwHKfiUI0...|  51.131.29.87|2009-04-13-08-05|
|   1239610365000|874NBpGmxNFfxEPKM...|xSvE4XtGbdtXPF2Lb...|cartoonnetwork.com|
Mozilla/5.0 (Maci...|eWDEVVUphlnRa273j...| 22.91.173.232|2009-04-13-08-05|
|   1239610348000|X8gISpUTSqh1A5reS...|TrFblGT99AgE75vuj...|       corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...|   55.35.44.79|2009-04-13-08-05|
|   1239610743000|kbKreLWB6QVueFrDm...|kVnxx9Ie2i3OLTxFj...|       corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...|   55.35.44.79|2009-04-13-08-05|
|   1239610812000|9lxOSRpEi3bmEeTCu...|1B2sff99AEIwSuLVV...|       corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...|   55.35.44.79|2009-04-13-08-05|
|   1239610876000|lijjmCf2kuxfBTnjL...|AjvufgUtakUFcsIM9...|       corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...|   55.35.44.79|2009-04-13-08-05|
|   1239610941000|t8t8trgjNRPIlmxuD...|agu2u2TCdqWP08rAA...|       corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...|   55.35.44.79|2009-04-13-08-05|
|   1239610490000|OGRLPVNGxiGgrCmWL...|mJg2raBUpPrC8OlUm...|       corriere.it|
Mozilla/4.0 (comp...|r2k96t1CNjSU9fJKN...|   71.124.66.3|2009-04-13-08-05|
|   1239610556000|OnJID12x0RXKPUgrD...|P7Pm2mPdW6wO8KA3R...|       corriere.it|
Mozilla/4.0 (comp...|r2k96t1CNjSU9fJKN...|   71.124.66.3|2009-04-13-08-05|
|   1239610373000|WflsvKIgOqfIE5KwR...|TJHd1VBspNcua0XPn...|       corriere.it|
Mozilla/5.0 (Maci...|fj2L1ILTFGMfhdrt3...| 75.117.56.155|2009-04-13-08-05|
|   1239610768000|4MJR0XxiVCU1ueXKV...|1OhGWmbvKf8ajoU8a...|       corriere.it|
Mozilla/5.0 (Maci...|fj2L1ILTFGMfhdrt3...| 75.117.56.155|2009-04-13-08-05|
|   1239610832000|gWIrpDiN57i3sHatv...|RNL4C7xPi3tdar2Uc...|       corriere.it|
Mozilla/5.0 (Maci...|fj2L1ILTFGMfhdrt3...| 75.117.56.155|2009-04-13-08-05|
|   1239610789000|pTne9k62kJ14QViXI...|RVxJVIQousjxUVI3r...|         pixnet.net|
Mozilla/5.0 (Maci...|1bGOKiBD2xmui9OkF...| 33.176.101.80|2009-04-13-08-05|
+----------------+--------------------+--------------------+------------------
+--------------------+--------------------+--------------+----------------+
only showing top 20 rows
```

```
scala>
```

Example: Create an HCatalog table and write to it using Pig

You can create an HCatalog table and use Apache Pig to write to it by way of HCatStorer using a data source in Amazon S3. HCatalog requires that you disable direct write, or the operation fails silently. Set both the `mapred.output.direct.NativeS3FileSystem` and the `mapred.output.direct.EmrFileSystem` configurations to `false` either using the `mapred-site` classification, or manually from within the Grunt shell. The following example shows a table created using the HCat CLI, followed by commands executed in the Grunt shell to populate the table from a sample data file in Amazon S3.

To run this example, connect to the master node using SSH.

Create an HCatalog script file, `wikicount.q`, with the following contents, which creates an HCatalog table named `wikicount`.

```
CREATE EXTERNAL TABLE IF NOT EXISTS wikicount(
col1 string,
col2 bigint
)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\001'
STORED AS ORC
LOCATION 's3://MyBucket/hcat/wikicount';
```

Use an HCat CLI command to execute the script from the file.

```
hcat -f wikicount.q
```

Next, start the Grunt shell with the `-useHCatalog` option, set configurations to disable direct write, load data from an S3 location, and then write the results to the wikicount table.

```
pig -useHCatalog
SET mapred.output.direct.NativeS3FileSystem false;
SET mapred.output.direct.EmrFileSystem false;
A = LOAD 's3://support.elasticmapreduce/training/datasets/wikistats_tiny/' USING
 PigStorage(' ') AS (Site:chararray, page:chararray, views:int, total_bytes:long);
B = GROUP A BY Site;
```

```
C = FOREACH B GENERATE group as col1, COUNT(A) as col2;
STORE C INTO 'wikicount' USING org.apache.hive.hcatalog.pig.HCatStorer();
```

HCatalog release history

The following table lists the version of HCatalog included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in Amazon EMR 7.x release versions, Amazon EMR 6.x release versions, or Amazon EMR 5.x release versions.

HCatalog version information

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-7.2.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.36.2	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
		timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-7.1.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-7.0.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-6.15.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.14.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-6.13.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.12.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-6.11.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.11.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-6.10.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.10.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-6.9.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.9.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-6.8.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.8.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-6.7.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.36.1	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.36.0	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.6.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.35.0	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.5.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-6.4.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.3.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-6.3.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.2.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-6.2.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.1.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-6.1.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.0.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-6.0.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.34.0	2.3.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.33.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.33.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.32.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.32.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.31.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.31.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.30.2	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.30.1	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.30.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.29.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.28.1	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.28.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.27.1	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.27.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.26.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.25.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.24.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.24.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.23.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.23.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.22.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.21.2	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.21.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.21.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.20.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.20.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.19.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.19.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.18.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.18.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.17.2	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.17.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.17.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.16.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.16.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.15.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.15.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.14.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.14.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.14.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.13.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.13.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.12.3	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.12.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.12.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.12.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.11.4	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.11.3	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.11.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.11.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.11.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.10.1	2.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.10.0	2.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.9.1	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.9.0	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.8.3	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.8.2	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.8.1	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.8.0	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.7.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.7.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.6.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.6.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.5.4	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.5.3	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.5.2	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.5.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.5.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.4.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.4.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.3.2	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.3.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.3.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.2.3	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.2.2	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.2.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.2.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.1.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.1.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.0.3	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.0.2	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.0.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-5.0.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.9.6	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-4.9.5	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.9.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-4.9.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.9.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-4.9.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.8.5	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-4.8.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.8.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-4.8.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.8.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-4.8.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.7.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-4.7.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.7.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-4.7.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.7.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-4.6.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, hive-metastore-server, mysql-server
emr-4.6.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, hive-metastore-server, mysql-server

Amazon EMR Release label	HCatalog Version	Components installed with HCatalog
emr-4.5.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, hive-metastore-server, mysql-server
emr-4.4.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, hive-metastore-server, mysql-server

Apache Hive

Hive is an open-source, data warehouse, and analytic package that runs on top of a Hadoop cluster. Hive scripts use an SQL-like language called Hive QL (query language) that abstracts programming models and supports typical data warehouse interactions. Hive enables you to avoid the complexities of writing Tez jobs based on directed acyclic graphs (DAGs) or MapReduce programs in a lower level computer language, such as Java.

Hive extends the SQL paradigm by including serialization formats. You can also customize query processing by creating table schema that match your data, without touching the data itself. While SQL only supports primitive value types, such as dates, numbers, and strings), Hive table values are structured elements, such as JSON objects, any user-defined data type, or any function written in Java.

For more information about Hive, see http://hive.apache.org/.

The following table lists the version of Hive included in the latest release of the Amazon EMR 7.x series, along with the components that Amazon EMR installs with Hive.

For the version of components installed with Hive in this release, see Release 7.2.0 Component Versions.

Hive version information for emr-7.2.0

Amazon EMR Release Label	Hive Version	Components Installed With Hive
emr-7.2.0	Hive 3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client,

Amazon EMR Release Label	Hive Version	Components Installed With Hive
		hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

The following table lists the version of Hive included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with Hive.

For the version of components installed with Hive in this release, see Release 6.15.0 Component Versions.

Hive version information for emr-6.15.0

Amazon EMR Release Label	Hive Version	Components Installed With Hive
emr-6.15.0	Hive 3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

The following table lists the version of Hive included in the latest release of the Amazon EMR 5.x series, along with the components that Amazon EMR installs with Hive.

For the version of components installed with Hive in this release, see Release 5.36.2 Component Versions.

Hive version information for emr-5.36.2

Amazon EMR Release Label	Hive Version	Components Installed With Hive
emr-5.36.2	Hive 2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Beginning with Amazon EMR 5.18.0, you can use the Amazon EMR artifact repository to build your job code against the exact versions of libraries and dependencies that are available with specific Amazon EMR releases. For more information, see Checking dependencies using the Amazon EMR artifact repository.

Topics

- Using Hive Live Long and Process (LLAP)
- Encryption in Hive
- Hive release history

Differences and considerations for Hive on Amazon EMR

Differences between Apache Hive on Amazon EMR and Apache Hive

This section describes the differences between Hive on Amazon EMR and the default versions of Hive available at http://svn.apache.org/viewvc/hive/branches/.

Hive authorization

Amazon EMR supports Hive authorization for HDFS but not for EMRFS and Amazon S3. Amazon EMR clusters run with authorization disabled by default.

Hive file merge behavior with Amazon S3

Apache Hive merges small files at the end of a map-only job if `hive.merge.mapfiles` is true and the merge is triggered only if the average output size of the job is less than the `hive.merge.smallfiles.avgsize` setting. Amazon EMR Hive has exactly the same behavior if the final output path is in HDFS. If the output path is in Amazon S3, the `hive.merge.smallfiles.avgsize` parameter is ignored. In that situation, the merge task is always triggered if `hive.merge.mapfiles` is set to `true`.

ACID transactions and Amazon S3

Amazon EMR 6.1.0 and later supports Hive ACID (Atomicity, Consistency, Isolation, Durability) transactions so it complies with the ACID properties of a database. With this feature, you can run INSERT, UPDATE, DELETE, and MERGE operations in Hive managed tables with data in Amazon Simple Storage Service (Amazon S3).

Hive Live Long and Process (LLAP)

LLAP functionality added in version 2.0 of default Apache Hive is not supported in Hive 2.1.0 on Amazon EMR release 5.0.

Amazon EMR version 6.0.0 and later supports the Live Long and Process (LLAP) functionality for Hive. For more information, see Using Hive LLAP.

Differences in Hive between Amazon EMR release version 4.x and 5.x

This section covers differences to consider before you migrate a Hive implementation from Hive version 1.0.0 on Amazon EMR release 4.x to Hive 2.x on Amazon EMR release 5.x.

Operational differences and considerations

- **Support added for ACID (atomicity, consistency, isolation, and durability) transactions:** This difference between Hive 1.0.0 on Amazon EMR 4.x and default Apache Hive has been eliminated.
- **Direct writes to Amazon S3 eliminated:** This difference between Hive 1.0.0 on Amazon EMR and the default Apache Hive has been eliminated. Hive 2.1.0 on Amazon EMR release 5.x now creates, reads from, and writes to temporary files stored in Amazon S3. As a result, to read from and write to the same table you no longer have to create a temporary table in the cluster's local HDFS file system as a workaround. If you use versioned buckets, be sure to manage these temporary files as described below.
- **Manage temp files when using Amazon S3 versioned buckets:** When you run Hive queries where the destination of generated data is Amazon S3, many temporary files and directories are created. This is new behavior as described earlier. If you use versioned S3 buckets, these temp files clutter Amazon S3 and incur cost if they're not deleted. Adjust your lifecycle rules so that data with a `/_tmp` prefix is deleted after a short period, such as five days. See Specifying a lifecycle configuration for more information.
- **Log4j updated to log4j 2:** If you use log4j, you may need to change your logging configuration because of this upgrade. See Apache log4j 2 for details.

Performance differences and considerations

- **Performance differences with Tez:** With Amazon EMR release 5.x , Tez is the default execution engine for Hive instead of MapReduce. Tez provides improved performance for most workflows.
- **Tables with many partitions:** Queries that generate a large number of dynamic partitions may fail, and queries that select from tables with many partitions may take longer than expected to execute. For example, a select from 100,000 partitions may take 10 minutes or more.

Additional features of Hive on Amazon EMR

Amazon EMR extends Hive with new features that support Hive integration with other AWS services, such as the ability to read from and write to Amazon Simple Storage Service (Amazon S3) and DynamoDB.

Variables in Hive

You can include variables in your scripts by using the dollar sign and curly braces.

```
add jar ${LIB}/jsonserde.jar
```

You pass the values of these variables to Hive on the command line using the `-d` parameter, as in the following example:

```
-d LIB=s3://elasticmapreduce/samples/hive-ads/lib
```

You can also pass the values into steps that execute Hive scripts.

To pass variable values into Hive steps using the console

1. Open the Amazon EMR console at https://console.aws.amazon.com/emr.
2. Choose **Create cluster**.
3. In the **Steps** section, for **Add Step**, choose **Hive Program** from the list and **Configure and add**.
4. In the **Add Step** dialog, specify the parameters using the following table as a guide, and then choose **Add**.

Field	Action
Script S3 location*	Specify the URI where your script resides in Amazon S3. The value must be in the form *BucketName* /*path*/*ScriptName* . For example: `s3://elasticmapreduce/samples/hive-ads/libs/response-time-stats.q` .

Field	Action
Input S3 location	Optionally, specify the URI where your input files reside in Amazon S3. The value must be in the form *BucketName* /*path*/. If specified , this will be passed to the Hive script as a parameter named INPUT. For example: `s3://elasticmapreduce/samples/hive-ads/tables/` .
Output S3 location	Optionally, specify the URI where you want the output stored in Amazon S3. The value must be in the form *BucketName* /*path*. If specified, this will be passed to the Hive script as a parameter named OUTPUT. For example: `s3://mybucket/hive-ads/output/` .
Arguments	Optionally, enter a list of arguments (space-separated strings) to pass to Hive. If you defined a path variable in your Hive script named ${SAMPLE} , for example: `CREATE EXTERNAL TABLE logs (requestBeginTime STRING, requestEndTime STRING, hostname STRING) PARTITIONED BY (dt STRING) \ ROW FORMAT serde 'com.amazon.elasticmapreduce.JsonSerde' WITH SERDEPROPERTIES ( 'paths'='requestBeginTime, requestEndTime, hostname' ) LOCATION '${SAMPLE}/tables/impressions';` To pass a value for the variable, type the following in the **Arguments** window: `-d SAMPLE=s3://elasticmapreduce/samples/hive-ads/` .
Action on Failure	This determines what the cluster does in response to any errors. The possible values for this setting are: • **Terminate cluster**: If the step fails, terminate the cluster. If the cluster has termination protection enabled AND keep alive enabled, it will not terminate. • **Cancel and wait**: If the step fails, cancel the remaining steps. If the cluster has keep alive enabled, the cluster will not terminate. • **Continue**: If the step fails, continue to the next step.

5. Select values as necessary and choose **Create cluster**.

To pass variable values into Hive steps using the AWS CLI

To pass variable values into Hive steps using the AWS CLI, use the `--steps` parameter and include an arguments list.

- **Note**

  Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.2.0 \
--applications Name=Hive Name=Pig --use-default-roles --ec2-attributes
 KeyName=myKey --instance-type m5.xlarge --instance-count 3 \
--steps Type=Hive,Name="Hive Program",ActionOnFailure=CONTINUE,Args=[-f,s3://
elasticmapreduce/samples/hive-ads/libs/response-time-stats.q,-d,INPUT=s3://
elasticmapreduce/samples/hive-ads/tables,-d,OUTPUT=s3://mybucket/hive-ads/output/,-
d,SAMPLE=s3://elasticmapreduce/samples/hive-ads/]
```

  For more information on using Amazon EMR commands in the AWS CLI, see https://docs.aws.amazon.com/cli/latest/reference/emr.

To pass variable values into Hive steps using the Java SDK

- The following example demonstrates how to pass variables into steps using the SDK. For more information, see Class StepFactory in the *AWS SDK for Java API Reference*.

```
StepFactory stepFactory = new StepFactory();

   StepConfig runHive = new StepConfig()
     .withName("Run Hive Script")
     .withActionOnFailure("TERMINATE_JOB_FLOW")
     .withHadoopJarStep(stepFactory.newRunHiveScriptStep(“s3://mybucket/script.q”,
      Lists.newArrayList(“-d”,”LIB= s3://elasticmapreduce/samples/hive-ads/lib”));
```

Amazon EMR Hive queries to accommodate partial DynamoDB schemas

Amazon EMR Hive provides maximum flexibility when querying DynamoDB tables by allowing you to specify a subset of columns on which you can filter data, rather than requiring your query to include all columns. This partial schema query technique is effective when you have a sparse database schema and want to filter records based on a few columns, such as filtering on time stamps.

The following example shows how to use a Hive query to:

- Create a DynamoDB table.
- Select a subset of items (rows) in DynamoDB and further narrow the data to certain columns.
- Copy the resulting data to Amazon S3.

```
DROP TABLE dynamodb;
DROP TABLE s3;

CREATE EXTERNAL TABLE dynamodb(hashKey STRING, recordTimeStamp BIGINT, fullColumn
 map<String, String>)
    STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
    TBLPROPERTIES (
     "dynamodb.table.name" = "myTable",
     "dynamodb.throughput.read.percent" = ".1000",
     "dynamodb.column.mapping" = "hashKey:HashKey,recordTimeStamp:RangeKey");

CREATE EXTERNAL TABLE s3(map<String, String>)
     ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
     LOCATION 's3://bucketname/path/subpath/';

INSERT OVERWRITE TABLE s3 SELECT item fullColumn FROM dynamodb WHERE recordTimeStamp <
 "2012-01-01";
```

The following table shows the query syntax for selecting any combination of items from DynamoDB.

Query example	Result description
SELECT * FROM *table_name* ;	Selects all items (rows) from a given table and includes data from all columns available for those items.
SELECT * FROM *table_name* WHERE *field_name* =*value*;	Selects some items (rows) from a given table and includes data from all columns available for those items.
SELECT *column1_name* , *column2_name* , *column3_name* FROM *table_name* ;	Selects all items (rows) from a given table and includes data from some columns available for those items.
SELECT *column1_name* , *column2_name* , *column3_name* FROM *table_name* WHERE *field_name* =*value*;	Selects some items (rows) from a given table and includes data from some columns available for those items.

Copy data between DynamoDB tables in different AWS Regions

Amazon EMR Hive provides a `dynamodb.region` property you can set per DynamoDB table. When `dynamodb.region` is set differently on two tables, any data you copy between the tables automatically occurs between the specified regions.

The following example shows you how to create a DynamoDB table with a Hive script that sets the `dynamodb.region` property:

> **Note**
>
> Per-table region properties override the global Hive properties.

```
CREATE EXTERNAL TABLE dynamodb(hashKey STRING, recordTimeStamp BIGINT, map<String,
 String> fullColumn)
    STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
    TBLPROPERTIES (
     "dynamodb.table.name" = "myTable",
     "dynamodb.region" = "eu-west-1",
```

```
     "dynamodb.throughput.read.percent" = ".1000",
     "dynamodb.column.mapping" = "hashKey:HashKey,recordTimeStamp:RangeKey");
```

Set DynamoDB throughput values per table

Amazon EMR Hive enables you to set the DynamoDB readThroughputPercent and writeThroughputPercent settings on a per table basis in the table definition. The following Amazon EMR Hive script shows how to set the throughput values. For more information about DynamoDB throughput values, see Specifying read and write requirements for tables.

```
CREATE EXTERNAL TABLE dynamodb(hashKey STRING, recordTimeStamp BIGINT, map<String,
 String> fullColumn)
    STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
    TBLPROPERTIES (
     "dynamodb.table.name" = "myTable",
     "dynamodb.throughput.read.percent" = ".4",
     "dynamodb.throughput.write.percent" = "1.0",
     "dynamodb.column.mapping" = "hashKey:HashKey,recordTimeStamp:RangeKey");
```

Configuring an external metastore for Hive

By default, Hive records metastore information in a MySQL database on the primary node's file system. The metastore contains a description of the table and the underlying data on which it is built, including the partition names, data types, and so on. When a cluster terminates, all cluster nodes shut down, including the primary node. When this happens, local data is lost because node file systems use ephemeral storage. If you need the metastore to persist, you must create an *external metastore* that exists outside the cluster.

You have two options for an external metastore:

- AWS Glue Data Catalog (Amazon EMR release 5.8.0 or later only).

  For more information, see Using the AWS Glue Data Catalog as the metastore for Hive.
- Amazon RDS or Amazon Aurora.

  For more information, see Using an external MySQL database or Amazon Aurora.

Note

If you're using Hive 3 and encounter too many connections to Hive metastore, configure the parameter `datanucleus.connectionPool.maxPoolSize` to have a smaller value or increase the number of connection the database server can handle. The increased number of connections is due to the way Hive computes the maximum number of JDBC connections. To calculate the optimal value for performance, see Hive Configuration Properties.

Using the AWS Glue Data Catalog as the metastore for Hive

Using Amazon EMR release 5.8.0 or later, you can configure Hive to use the AWS Glue Data Catalog as its metastore. We recommend this configuration when you require a persistent metastore or a metastore shared by different clusters, services, applications, or AWS accounts.

AWS Glue is a fully managed extract, transform, and load (ETL) service that makes it simple and cost-effective to categorize your data, clean it, enrich it, and move it reliably between various data stores. The AWS Glue Data Catalog provides a unified metadata repository across a variety of data sources and data formats, integrating with Amazon EMR as well as Amazon RDS, Amazon Redshift, Redshift Spectrum, Athena, and any application compatible with the Apache Hive metastore. AWS Glue crawlers can automatically infer schema from source data in Amazon S3 and store the associated metadata in the Data Catalog. For more information about the Data Catalog, see Populating the AWS Glue Data Catalog in the *AWS Glue Developer Guide*.

Separate charges apply for AWS Glue. There is a monthly rate for storing and accessing the metadata in the Data Catalog, an hourly rate billed per minute for AWS Glue ETL jobs and crawler runtime, and an hourly rate billed per minute for each provisioned development endpoint. The Data Catalog allows you to store up to a million objects at no charge. If you store more than a million objects, you are charged USD$1 for each 100,000 objects over a million. An object in the Data Catalog is a table, partition, or database. For more information, see Glue Pricing.

Important

If you created tables using Amazon Athena or Amazon Redshift Spectrum before August 14, 2017, databases and tables are stored in an Athena-managed catalog, which is separate from the AWS Glue Data Catalog. To integrate Amazon EMR with these tables, you must upgrade to the AWS Glue Data Catalog. For more information, see Upgrading to the AWS Glue Data Catalog in the *Amazon Athena User Guide*.

Specifying AWS Glue Data Catalog as the metastore

You can specify the AWS Glue Data Catalog as the metastore using the AWS Management Console, AWS CLI, or Amazon EMR API. When you use the CLI or API, you use the configuration classification for Hive to specify the Data Catalog. In addition, with Amazon EMR 5.16.0 and later, you can use the configuration classification to specify a Data Catalog in a different AWS account. When you use the console, you can specify the Data Catalog using **Advanced Options** or **Quick Options**.

Console

To specify AWS Glue Data Catalog as the Hive metastore with the console

1. Sign in to the AWS Management Console, and open the Amazon EMR console at https://console.aws.amazon.com/emr.
2. Under **EMR on EC2** in the left navigation pane, choose **Clusters**, and then choose **Create cluster**.
3. Under **Application bundle**, choose **Core Hadoop**, **HBase**, or **Custom**. If you customize your cluster, make sure that you select Hive or HCatalog as one of your applications.
4. Under **AWS Glue Data Catalog settings**, select the **Use for Hive table metadata** check box.
5. Choose any other options that apply to your cluster.
6. To launch your cluster, choose **Create cluster**.

CLI

To specify the AWS Glue Data Catalog as the Hive metastore with the AWS CLI

For more information about specifying a configuration classification using the AWS CLI and EMR API, see Configure applications.

- Specify the value for `hive.metastore.client.factory.class` using the `hive-site` configuration classification as shown in the following example:

```
[
  {
```

```
    "Classification": "hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
 "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory"
    }
  }
]
```

On EMR release versions 5.28.0, 5.28.1, 5.29.0, or 6.x, if you're creating a cluster using the AWS Glue Data Catalog as the metastore, set the `hive.metastore.schema.verification` to `false`. This prevents Hive and HCatalog from validating the metastore schema against MySQL. Without this configuration, the primary instance group will become suspended after reconfiguration on Hive or HCatalog.

```
[
  {
    "Classification": "hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
 "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory",
      "hive.metastore.schema.verification": "false"
    }
  }
]
```

If you already have a cluster on EMR release version 5.28.0, 5.28.1, or 5.29.0, you can set the primary instance group `hive.metastore.schema.verification` to `false` with following information:

```
    Classification = hive-site
    Property       = hive.metastore.schema.verification
    Value          = false
```

To specify a Data Catalog in a different AWS account, add the `hive.metastore.glue.catalogid` property as shown in the following example. Replace *`acct-id`* with the AWS account of the Data Catalog.

```
[
  {
    "Classification": "hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
 "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory",
      "hive.metastore.schema.verification": "false",
      "hive.metastore.glue.catalogid": "acct-id"
    }
  }
]
```

IAM permissions

The EC2 instance profile for a cluster must have IAM permissions for AWS Glue actions. In addition, if you enable encryption for AWS Glue Data Catalog objects, the role must also be allowed to encrypt, decrypt and generate the AWS KMS key used for encryption.

Permissions for AWS Glue actions

If you use the default EC2 instance profile for Amazon EMR, no action is required. The `AmazonElasticMapReduceforEC2Role` managed policy that is attached to the `EMR_EC2_DefaultRole` allows all necessary AWS Glue actions. However, if you specify a custom EC2 instance profile and permissions, you must configure the appropriate AWS Glue actions. Use the `AmazonElasticMapReduceforEC2Role` managed policy as a starting point. For more information, see Service role for cluster EC2 instances (EC2 instance profile) in the *Amazon EMR Management Guide*.

Permissions for encrypting and decrypting AWS Glue Data Catalog

Your instance profile needs permission to encrypt and decrypt data using your key. You do *not* need to configure these permissions if both of the following statements apply:

- You enable encryption for AWS Glue Data Catalog objects using managed keys for AWS Glue.
- You use a cluster that's in the same AWS account as the AWS Glue Data Catalog.

Otherwise, you must add the following statement to the permissions policy attached to your EC2 instance profile.

```
[
    {
        "Version": "2012-10-17",
        "Statement": [
            {
                "Effect": "Allow",
                "Action": [
                    "kms:Decrypt",
                    "kms:Encrypt",
                    "kms:GenerateDataKey"
                ],
                "Resource": "arn:aws:kms:region:acct-
id:key/12345678-1234-1234-1234-123456789012"
            }
        ]
    }
]
```

For more information about AWS Glue Data Catalog encryption, see Encrypting your data catalog in the *AWS Glue Developer Guide*.

Resource-based permissions

If you use AWS Glue in conjunction with Hive, Spark, or Presto in Amazon EMR, AWS Glue supports resource-based policies to control access to Data Catalog resources. These resources include databases, tables, connections, and user-defined functions. For more information, see AWS Glue Resource Policies in the *AWS Glue Developer Guide*.

When using resource-based policies to limit access to AWS Glue from within Amazon EMR, the principal that you specify in the permissions policy must be the role ARN associated with the EC2 instance profile that is specified when a cluster is created. For example, for a resource-based policy attached to a catalog, you can specify the role ARN for the default service role for cluster EC2 instances, *EMR_EC2_DefaultRole* as the `Principal`, using the format shown in the following example:

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

The *acct-id* can be different from the AWS Glue account ID. This enables access from EMR clusters in different accounts. You can specify multiple principals, each from a different account.

Considerations when using AWS Glue Data Catalog

Consider the following items when using the AWS Glue Data Catalog as the metastore with Hive:

- Adding auxiliary JARs using the Hive shell is not supported. As a workaround, use the `hive-site` configuration classification to set the `hive.aux.jars.path` property, which adds auxiliary JARs into the Hive classpath.
- Hive transactions are not supported.
- Renaming tables from within AWS Glue is not supported.
- When you create a Hive table without specifying a `LOCATION`, the table data is stored in the location specified by the `hive.metastore.warehouse.dir` property. By default, this is a location in HDFS. If another cluster needs to access the table, it fails unless it has adequate permissions to the cluster that created the table. Furthermore, because HDFS storage is transient, if the cluster terminates, the table data is lost, and the table must be recreated. We recommend that you specify a `LOCATION` in Amazon S3 when you create a Hive table using AWS Glue. Alternatively, you can use the `hive-site` configuration classification to specify a location in Amazon S3 for `hive.metastore.warehouse.dir`, which applies to all Hive tables. If a table is created in an HDFS location and the cluster that created it is still running, you can update the table location to Amazon S3 from within AWS Glue. For more information, see Working with Tables on the AWS Glue Console in the *AWS Glue Developer Guide*.
- Partition values containing quotes and apostrophes are not supported, for example, `PARTITION (owner="Doe's").`
- Column statistics are supported for emr-5.31.0 and later.
- Using Hive authorization is not supported. As an alternative, consider using AWS Glue Resource-Based Policies. For more information, see Use Resource-Based Policies for Amazon EMR Access to AWS Glue Data Catalog.
- Hive constraints are not supported.
- Cost-based Optimization in Hive is not supported.
- Setting `hive.metastore.partition.inherit.table.properties` is not supported.
- Using the following metastore constants is not supported: `BUCKET_COUNT, BUCKET_FIELD_NAME, DDL_TIME, FIELD_TO_DIMENSION, FILE_INPUT_FORMAT, FILE_OUTPUT_FORMAT, HIVE_FILTER_FIELD_LAST_ACCESS, HIVE_FILTER_FIELD_OWNER, HIVE_FILTER_FIELD_PARAMS, IS_ARCHIVED, META_TABLE_COLUMNS, META_TABLE_COLUMN_TYPES, META_TABLE_DB,`

`META_TABLE_LOCATION`, `META_TABLE_NAME`, `META_TABLE_PARTITION_COLUMNS`, `META_TABLE_SERDE`, `META_TABLE_STORAGE`, `ORIGINAL_LOCATION`.

- When you use a predicate expression, explicit values must be on the right side of the comparison operator, or queries might fail.
  - **Correct**: `SELECT * FROM mytable WHERE time > 11`
  - **Incorrect**: `SELECT * FROM mytable WHERE 11 > time`
- Amazon EMR versions 5.32.0 and 6.3.0 and later support using user-defined functions (UDFs) in predicate expressions. When using earlier versions, your queries may fail because of the way Hive tries to optimize query execution.
- Temporary tables are not supported.
- We recommend creating tables using applications through Amazon EMR rather than creating them directly using AWS Glue. Creating a table through AWS Glue may cause required fields to be missing and cause query exceptions.
- In EMR 5.20.0 or later, parallel partition pruning is enabled automatically for Spark and Hive when AWS Glue Data Catalog is used as the metastore. This change significantly reduces query planning time by executing multiple requests in parallel to retrieve partitions. The total number of segments that can be executed concurrently range between 1 and 10. The default value is 5, which is a recommended setting. You can change it by specifying the property `aws.glue.partition.num.segments` in `hive-site` configuration classification. If throttling occurs, you can turn off the feature by changing the value to 1. For more information, see AWS Glue Segment Structure.

Using an external MySQL database or Amazon Aurora

To use an external MySQL database or Amazon Aurora as your Hive metastore, you override the default configuration values for the metastore in Hive to specify the external database location, either on an Amazon RDS MySQL instance or an Amazon Aurora PostgreSQLinstance.

> **Note**
>
> Hive neither supports nor prevents concurrent write access to metastore tables. If you share metastore information between two clusters, you must ensure that you do not write to the same metastore table concurrently, unless you are writing to different partitions of the same metastore table.

The following procedure shows you how to override the default configuration values for the Hive metastore location and start a cluster using the reconfigured metastore location.

To create a metastore located outside of the EMR cluster

1. Create a MySQL or Aurora PostgreSQL database. If you use PostgreSQL, you must configure it after you've provisioned your cluster. Only MySQL is supported at cluster creation. For information about the differences between Aurora MySQL and Aurora PostgreSQL, see Overview of Amazon Aurora MySQL and Working with Amazon Aurora PostgreSQL. For information about how to create an Amazon RDS database in general, see https://aws.amazon.com/rds/.
2. Modify your security groups to allow JDBC connections between your database and the **ElasticMapReduce-Master** security group. For information about how to modify your security groups for access, see Working with Amazon EMR-managed security groups.
3. Set JDBC configuration values in `hive-site.xml`:

> **Important**
>
> If you supply sensitive information, such as passwords, to the Amazon EMR configuration API, this information is displayed for those accounts that have sufficient permissions. If you are concerned that this information could be displayed to other users, create the cluster with an administrative account and limit other users (IAM users or those with delegated credentials) to accessing services on the cluster by creating a role which explicitly denies permissions to the `elasticmapreduce:DescribeCluster` API key.

   a. Create a configuration file called `hiveConfiguration.json` containing edits to `hive-site.xml` as shown in the following example.

      Replace *`hostname`* with the DNS address of your Amazon RDS instance running the database, and *`username`* and *`password`* with the credentials for your database. For more information about connecting to MySQL and Aurora database instances, see Connecting to a DB instance running the MySQL database engine and Connecting to an Athena DB cluster in the *Amazon RDS User Guide*. `javax.jdo.option.ConnectionURL` is the JDBC connect string for a JDBC metastore. `javax.jdo.option.ConnectionDriverName` is the driver class name for a JDBC metastore.

The MySQL JDBC drivers are installed by Amazon EMR.

The value property can not contain any spaces or carriage returns. It should appear all on one line.

```
[
    {
      "Classification": "hive-site",
      "Properties": {
        "javax.jdo.option.ConnectionURL": "jdbc:mysql://hostname:3306/hive?
createDatabaseIfNotExist=true",
        "javax.jdo.option.ConnectionDriverName": "org.mariadb.jdbc.Driver",
        "javax.jdo.option.ConnectionUserName": "username",
        "javax.jdo.option.ConnectionPassword": "password"
      }
    }
  ]
```

b. Reference the `hiveConfiguration.json` file when you create the cluster as shown in the following AWS CLI command. In this command, the file is stored locally, you can also upload the file to Amazon S3 and reference it there, for example, `s3://DOC-EXAMPLE-BUCKET/hiveConfiguration.json`.

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge --
instance-count 2 \
--applications Name=Hive --configurations file://hiveConfiguration.json --use-
default-roles
```

4. Connect to the primary node of your cluster.

   For information about how to connect to the primary node, see [Connect to the primary node using SSH](#) in the *Amazon EMR Management Guide*.

5. Create your Hive tables specifying the location on Amazon S3 by entering a command similar to the following:

```
CREATE EXTERNAL TABLE IF NOT EXISTS table_name
(
key int,
value int
)
LOCATION s3://DOC-EXAMPLE-BUCKET/hdfs/
```

6. Add your Hive script to the running cluster.

Your Hive cluster runs using the metastore located in Amazon RDS. Launch all additional Hive clusters that share this metastore by specifying the metastore location.

Use the Hive JDBC driver

You can use popular business intelligence tools like Microsoft Excel, MicroStrategy, QlikView, and Tableau with Amazon EMR to explore and visualize your data. Many of these tools require Java Database Connectivity (JDBC) driver or an Open Database Connectivity (ODBC) driver. Amazon EMR supports both JDBC and ODBC connectivity.

The example below demonstrates using SQL Workbench/J as a SQL client to connect to a Hive cluster in Amazon EMR. For additional drivers, see Use business intelligence tools with Amazon EMR.

Before you install and work with SQL Workbench/J, download the driver package and install the driver. The drivers included in the package support the Hive versions available in Amazon EMR release versions 4.0 and later. For detailed release notes and documentation, see the PDF documentation included in the package.

- **The latest Hive JDBC driver package download**

  http://awssupportdatasvcs.com/bootstrap-actions/Simba/latest/

- **Older versions of the Hive JDBC driver**

  http://awssupportdatasvcs.com/bootstrap-actions/Simba/

To install and configure SQL Workbench

1. Download the SQL Workbench/J client for your operating system from http://www.sql-workbench.net/downloads.html.
2. Install SQL Workbench/J. For more information, see [Installing and starting SQL Workbench/J](#) in the SQL Workbench/J Manual User's Manual.
3. **Linux, Unix, Mac OS X users**: In a terminal session, create an SSH tunnel to the master node of your cluster using the following command. Replace *master-public-dns-name* with the public DNS name of the master node and *path-to-key-file* with the location and file name of your Amazon EC2 private key (`.pem`) file.

```
ssh -o ServerAliveInterval=10 -i path-to-key-file -N -L 10000:localhost:10000
 hadoop@master-public-dns-name
```

   Windows users: In a PuTTY session, create an SSH tunnel to the master node of your cluster (using local port forwarding) with `10000` for **Source port** and *master-public-dns-name*`:10000` for **Destination**. Replace *master-public-dns-name* with the public DNS name of the master node.

4. Add the JDBC driver to SQL Workbench.

   a. In the **Select Connection Profile** dialog box, click **Manage Drivers**.

   b. Click the **Create a new entry** (blank page) icon.

   c. In the **Name** field, type **`Hive JDBC`**.

   d. For **Library**, click the **Select the JAR file(s)** icon.

   e. Navigate to the location containing the extracted drivers. Select the drivers that are included in the JDBC driver package version that you downloaded, and click **Open**.

      For example, your JDBC driver package may include the following JARs.

```
hive_metastore.jar
hive_service.jar
HiveJDBC41.jar
libfb303-0.9.0.jar
libthrift-0.9.0.jar
log4j-1.2.14.jar
ql.jar
slf4j-api-1.5.11.jar
slf4j-log4j12-1.5.11.jar
```

```
TCLIServiceClient.jar
zookeeper-3.4.6.jar
```

f. In the **Please select one driver** dialog box, select `com.amazon.hive.jdbc41.HS2Driver`, **OK**.

5. When you return to the **Manage Drivers** dialog box, verify that the **Classname** field is populated and select **OK**.
6. When you return to the **Select Connection Profile** dialog box, verify that the **Driver** field is set to **Hive JDBC** and provide the following JDBC connection string in the **URL** field: `jdbc:hive2://localhost:10000/default`.
7. Select **OK** to connect. After the connection is complete, connection details appear at the top of the SQL Workbench/J window.

For more information about using Hive and the JDBC interface, see HiveClient and HiveJDBCInterface in Apache Hive documentation.

Improve Hive performance

Amazon EMR offers features to help optimize performance when using Hive to query, read and write data saved in Amazon S3.

S3 Select can improve query performance for CSV and JSON files in some applications by “pushing down” processing to Amazon S3.

The EMRFS S3 optimized committer is an alternative to the OutputCommitter class, that eliminates list and rename operations to improve performance when writing files Amazon S3 using EMRFS.

Topics

Enabling Hive EMRFS S3 optimized committer

The Hive EMRFS S3 Optimized Committer is an alternative way using which EMR Hive writes files for insert queries when using EMRFS. The Committer eliminates list and rename operations done on Amazon S3 and improves application's performance. The feature is available beginning with EMR 5.34 and EMR 6.5.

Enabling the committer

If you want to enable EMR Hive to use `HiveEMRFSOptimizedCommitter` to commit data as the default for all Hive managed and external tables, use the following `hive-site` configuration in EMR 6.5.0 or EMR 5.34.0 clusters.

```
[
   {
      "classification": "hive-site",
      "properties": {
         "hive.blobstore.use.output-committer": "true"
      }
   }
]
```

> **Note**
>
> Do not turn this feature on when `hive.exec.parallel` is set to `true`.

Limitations

The following basic restrictions apply to tags:

- Enabling Hive to merge small files automatically is not supported. The default Hive commit logic will be used even when the optimized committer is enabled.
- Hive ACID tables are not supported. The default Hive commit logic will be used even when the optimized committer is enabled.
- File naming nomenclature for files written is changed from Hive's <task_id>_<attempt_id>_<copy_n> to <task_id>_<attempt_id>_<copy_n>_<query_id>. For example, a file named

  `s3://warehouse/table/partition=1/000000_0` will be changed to `s3://warehouse/table/partition=1/000000_0-hadoop_20210714130459_ba7c23ec-5695-4947-9d98-8a40ef759222-1`. The `query_id` here is a combination of the username, time stamp, and UUID.

- When custom partitions are on different file systems (HDFS, S3), this feature is automatically disabled. The default Hive commit logic will be used when enabled.

Using S3 Select with Hive to improve performance

With Amazon EMR release version 5.18.0 and later, you can use S3 Select with Hive on Amazon EMR. S3 Select allows applications to retrieve only a subset of data from an object. For Amazon EMR, the computational work of filtering large datasets for processing is "pushed down" from the cluster to Amazon S3, which can improve performance in some applications and reduces the amount of data transferred between Amazon EMR and Amazon S3.

S3 Select is supported with Hive tables based on CSV and JSON files and by setting the `s3select.filter` configuration variable to `true` during your Hive session. For more information and examples, see Specifying S3 Select in your code.

Is S3 Select right for my application?

We recommend that you benchmark your applications with and without S3 Select to see if using it may be suitable for your application.

Use the following guidelines to determine if your application is a candidate for using S3 Select:

- Your query filters out more than half of the original dataset.
- Your query filter predicates use columns that have a data type supported by Amazon S3 Select. For more information, see Data types in the *Amazon Simple Storage Service User Guide*.
- Your network connection between Amazon S3 and the Amazon EMR cluster has good transfer speed and available bandwidth. Amazon S3 does not compress HTTP responses, so the response size is likely to increase for compressed input files.

Considerations and limitations

- Amazon S3 server-side encryption with customer-provided encryption keys (SSE-C) and client-side encryption are not supported.
- The `AllowQuotedRecordDelimiters` property is not supported. If this property is specified, the query fails.
- Only CSV and JSON files in UTF-8 format are supported. Multi-line CSVs and JSON are not supported.

- Only uncompressed or gzip or bzip2 files are supported.
- Comment characters in the last line are not supported.
- Empty lines at the end of a file are not processed.
- Hive on Amazon EMR supports the primitive data types that S3 Select supports. For more information, see Data types in the *Amazon Simple Storage Service User Guide*.

Specifying S3 Select in your code

To use S3 Select in your Hive table, create the table by specifying `com.amazonaws.emr.s3select.hive.S3SelectableTextInputFormat` as the `INPUTFORMAT` class name, and specify a value for the `s3select.format` property using the `TBLPROPERTIES` clause.

By default, S3 Select is disabled when you run queries. Enable S3 Select by setting `s3select.filter` to `true` in your Hive session as shown below. The examples below demonstrate how to specify S3 Select when creating a table from underlying CSV and JSON files and then querying the table using a simple select statement.

Example CREATE TABLE statement for CSV-based table

```
CREATE TABLE mys3selecttable (
col1 string,
col2 int,
col3 boolean
)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS
INPUTFORMAT
  'com.amazonaws.emr.s3select.hive.S3SelectableTextInputFormat'
OUTPUTFORMAT
  'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'
LOCATION 's3://path/to/mycsvfile/'
TBLPROPERTIES (
  "s3select.format" = "csv",
  "s3select.headerInfo" = "ignore"
);
```

Example CREATE TABLE statement for JSON-based table

```
CREATE TABLE mys3selecttable (
```

```
col1 string,
col2 int,
col3 boolean
)
ROW FORMAT SERDE 'org.apache.hive.hcatalog.data.JsonSerDe'
STORED AS
INPUTFORMAT
  'com.amazonaws.emr.s3select.hive.S3SelectableTextInputFormat'
OUTPUTFORMAT
  'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'
LOCATION 's3://path/to/json/'
TBLPROPERTIES (
  "s3select.format" = "json"
);
```

Example SELECT TABLE statement

```
SET s3select.filter=true;
SELECT * FROM mys3selecttable WHERE col2 > 10;
```

MSCK Optimization

Hive stores a list of partitions for each table in its metastore. However, when partitions are directly added to or removed from the file system, the Hive metastore is unaware of these changes. The MSCK command updates the partition metadata in the Hive metastore for partitions that were directly added to or removed from the file system. The syntax for the command is:

```
MSCK [REPAIR] TABLE table_name [ADD/DROP/SYNC PARTITIONS];
```

Hive implements this command as follows:

1. Hive retrieves all the partitions for the table from the metastore. From the list of partition paths that do not exist in the file system then creates a list of partitions to drop from the metastore.
2. Hive gathers the partition paths present in the file system, compares them with the list of partitions from the metastore, and generates a list of partitions that need to be added to the metastore.
3. Hive updates the metastore using ADD, DROP, or SYNC mode.

> **Note**
>
> When there are many partitions in the metastore, the step to check if a partition does not exist in the file system takes a long time to run because the file system's `exists` API call must be made for each partition.

In Amazon EMR 6.5.0, Hive introduced a flag called `hive.emr.optimize.msck.fs.check`. When enabled, this flag causes Hive to check for the presence of a partition from the list of partition paths from the file system that is generated in step 2 above instead of making file system API calls. In Amazon EMR 6.8.0, Hive enabled this optimization by default, eliminating the need to set the flag `hive.emr.optimize.msck.fs.check`.

Using Hive Live Long and Process (LLAP)

Amazon EMR 6.0.0 supports the Live Long and Process (LLAP) functionality for Hive. LLAP uses persistent daemons with intelligent in-memory caching to improve query performance compared to the previous default Tez container execution mode.

The Hive LLAP daemons are managed and run as a YARN Service. Since a YARN service can be considered a long-running YARN application, some of your cluster resources are dedicated to Hive LLAP and cannot be used for other workloads. For more information, see LLAP and YARN Service API.

Enable Hive LLAP on Amazon EMR

To enable Hive LLAP on Amazon EMR, supply the following configuration when you launch a cluster.

```
[
  {
    "Classification": "hive",
    "Properties": {
      "hive.llap.enabled": "true"
    }
  }
]
```

For more information, see Configuring applications.

By default, Amazon EMR allocates about 60 percent of cluster YARN resources to Hive LLAP daemons. You can configure the percentage of cluster YARN resource allocated to Hive LLAP and the number of task and core nodes to be considered for the Hive LLAP allocation.

For example, the following configuration starts Hive LLAP with three daemons on three task or core nodes and allocates 40 percent of the three core or task nodes' YARN resource to the Hive LLAP daemons.

```
[
  {
    "Classification": "hive",
    "Properties": {
      "hive.llap.enabled": "true",
      "hive.llap.percent-allocation": "0.4",
      "hive.llap.num-instances": "3"
    }
  }
]
```

You can use the following `hive-site` configurations in the classification API to override default LLAP resource settings.

Property	Description
hive.llap.daemon.yarn.container.mb	Total LLAP daemon container size (in MB)
hive.llap.daemon.memory.per.instance.mb	The total memory used by executors in the LLAP daemon container (in MB)
hive.llap.io.memory.size	Cache size for LLAP Input/Output
hive.llap.daemon.num.executors	Number of executors per LLAP daemon

Start Hive LLAP on your cluster manually

All dependencies and configurations used by LLAP are packaged into the LLAP tar archive as part of cluster startup. If LLAP is enabled using `"hive.llap.enabled": "true"`, we recommend that you use Amazon EMR reconfiguration to make configuration changes to LLAP.

Otherwise, for any manual changes to `hive-site.xml`, you must rebuild the LLAP tar archive by using the `hive --service llap` command, as the following example demonstrates.

```
# Define how many resources you want to allocate to Hive LLAP

LLAP_INSTANCES=<how many llap daemons to run on cluster>
LLAP_SIZE=<total container size per llap daemon>
LLAP_EXECUTORS=<number of executors per daemon>
LLAP_XMX=<Memory used by executors>
LLAP_CACHE=<Max cache size for IO allocator>

yarn app -enableFastLaunch

hive --service llap \
--instances $LLAP_INSTANCES \
--size ${LLAP_SIZE}m \
--executors $LLAP_EXECUTORS \
--xmx ${LLAP_XMX}m \
--cache ${LLAP_CACHE}m \
--name llap0 \
--auxhbase=false \
--startImmediately
```

Check Hive LLAP status

Use the following command to check the status of Hive LLAP through Hive.

```
hive --service llapstatus
```

Use the following command to check the status of Hive LLAP using YARN.

```
yarn app -status (name-of-llap-service)

# example:
```

```
yarn app -status llap0 | jq
```

Start or stop Hive LLAP

Since Hive LLAP runs as a persistent YARN service, you stop or restart the YARN service to stop or restart Hive LLAP. The following commands demonstrate this.

```
yarn app -stop llap0
yarn app -start llap0
```

Resize the number of Hive LLAP daemons

Use the following command to reduce the number of LLAP instances.

```
yarn app -flex llap0 -component llap -1
```

For more information, see Flex a component of a service.

Encryption in Hive

This section describes the encryption types Amazon EMR supports.

Parquet modular encryption in Hive

Parquet modular encryption provides columnar level access control and encryption to enhance privacy and data integrity for data stored in Parquet file format. This feature is available in Amazon EMR Hive starting with release 6.6.0.

Previously supported solutions for security and integrity, which include encrypting files or encrypting the storage layer, are described in Encryption Options in the Amazon EMR Management Guide. These solutions can be used for Parquet files, but leveraging the new features of the integrated Parquet encryption mechanism provides granular access to the column level, as well as improvements in performance and security. Learn more about this feature on the Apache github page Parquet Modular Encryption.

Users pass configurations to Parquet readers and writers using Hadoop configurations. The detailed configurations for users to configure readers and writers to enable encryption as well as toggle advanced features are documented at [PARQUET-1854: Properties-driven Interface to Parquet Encryption Management]

Usage examples

The following example covers creating and writing to a Hive table using AWS KMS for managing encryption keys.

1. Implement a KmsClient for the AWS KMS service as described in the document [PARQUET-1373: Encryption Key Management Tools]. The following sample shows an implementation snippet.

```
package org.apache.parquet.crypto.keytools;

import com.amazonaws.AmazonClientException;
import com.amazonaws.AmazonServiceException;
import com.amazonaws.regions.Regions;
import com.amazonaws.services.kms.AWSKMS;
import com.amazonaws.services.kms.AWSKMSClientBuilder;
import com.amazonaws.services.kms.model.DecryptRequest;
import com.amazonaws.services.kms.model.EncryptRequest;
import com.amazonaws.util.Base64;
import org.apache.hadoop.conf.Configuration;
import org.apache.parquet.crypto.KeyAccessDeniedException;
import org.apache.parquet.crypto.ParquetCryptoRuntimeException;
import org.apache.parquet.crypto.keytools.KmsClient;
import org.slf4j.Logger;
import org.slf4j.LoggerFactory;

import java.nio.ByteBuffer;
import java.nio.charset.Charset;
import java.nio.charset.StandardCharsets;

public class AwsKmsClient implements KmsClient {

    private static final AWSKMS AWSKMS_CLIENT = AWSKMSClientBuilder
            .standard()
            .withRegion(Regions.US_WEST_2)
            .build();
    public static final Logger LOG = LoggerFactory.getLogger(AwsKmsClient.class);

    private String kmsToken;
    private Configuration hadoopConfiguration;
```

```
    @Override
    public void initialize(Configuration configuration, String kmsInstanceID, String
 kmsInstanceURL, String accessToken) throws KeyAccessDeniedException {
        hadoopConfiguration = configuration;
        kmsToken = accessToken;

    }

    @Override
    public String wrapKey(byte[] keyBytes, String masterKeyIdentifier) throws
 KeyAccessDeniedException {
        String value = null;
        try {
            ByteBuffer plaintext = ByteBuffer.wrap(keyBytes);

            EncryptRequest req = new
 EncryptRequest().withKeyId(masterKeyIdentifier).withPlaintext(plaintext);
            ByteBuffer ciphertext = AWSKMS_CLIENT.encrypt(req).getCiphertextBlob();

            byte[] base64EncodedValue = Base64.encode(ciphertext.array());
            value = new String(base64EncodedValue, Charset.forName("UTF-8"));
        } catch (AmazonClientException ae) {
            throw new KeyAccessDeniedException(ae.getMessage());
        }
        return value;
    }

    @Override
    public byte[] unwrapKey(String wrappedKey, String masterKeyIdentifier) throws
 KeyAccessDeniedException {
        byte[] arr = null;
        try {
            ByteBuffer ciphertext  =
 ByteBuffer.wrap(Base64.decode(wrappedKey.getBytes(StandardCharsets.UTF_8)));
            DecryptRequest request = new
 DecryptRequest().withKeyId(masterKeyIdentifier).withCiphertextBlob(ciphertext);
            ByteBuffer decipheredtext =
 AWSKMS_CLIENT.decrypt(request).getPlaintext();
            arr = new byte[decipheredtext.remaining()];
            decipheredtext.get(arr);
        } catch (AmazonClientException ae) {
            throw new KeyAccessDeniedException(ae.getMessage());
        }
        return arr;
```

```
    }
}
```

2. Create your AWS KMS encryption keys for the footer as well the columns with your IAM roles having access as described in [Creating keys](#) in the *AWS Key Management Service Developer Guide*. The default IAM role is EMR_ECS_default.
3. On the Hive application on an Amazon EMR cluster, add the client above using the `ADD JAR` statement, as described in the [Apache Hive Resources documentation](#). The following is an example statement:

```
ADD JAR 's3://location-to-custom-jar';
```

   An alternative method is to add the JAR to the `auxlib` of Hive using a bootstrap action. The following is an example line to be added to the boostrap action:

```
aws s3 cp 's3://location-to-custom-jar' /usr/lib/hive/auxlib
```

4. Set the following configurations:

```
set
 parquet.crypto.factory.class=org.apache.parquet.crypto.keytools.PropertiesDrivenCryptoFacto
set
 parquet.encryption.kms.client.class=org.apache.parquet.crypto.keytools.AwsKmsClient;
```

5. Create a Hive table with Parquet format and specify the AWS KMS keys in SERDEPROPERTIES and insert some data to it:

```
CREATE TABLE my_table(name STRING, credit_card STRING)
ROW FORMAT SERDE 'org.apache.hadoop.hive.ql.io.parquet.serde.ParquetHiveSerDe’
WITH SERDEPROPERTIES (
  'parquet.encryption.column.key’=<aws-kms-key-id-for-column-1>: credit_card’,
  'parquet.encryption.footer.key’='<aws-kms-key-id-for-footer>’)
STORED AS parquet
LOCATION “s3://<bucket/<warehouse-location>/my_table”;

INSERT INTO my_table SELECT
java_method ('org.apache.commons.lang.RandomStringUtils','randomAlphabetic',5) as
 name,
java_method ('org.apache.commons.lang.RandomStringUtils','randomAlphabetic',10) as
 credit_card
```

```
from (select 1) x lateral view posexplode(split(space(100),' ')) pe as i,x;

select * from my_table;
```

6. Verify that when you create an external table at the same location with no access to AWS KMS keys (for example, IAM role access denied), you cannot read the data.

```
CREATE EXTERNAL TABLE ext_table (name STRING, credit_card STRING)
ROW FORMAT SERDE 'org.apache.hadoop.hive.ql.io.parquet.serde.ParquetHiveSerDe’
STORED AS parquet
LOCATION “s3://<bucket>/<warehouse-location>/my_table”;

SELECT * FROM ext_table;
```

7. The last statement should throw the following exception:

```
Failed with exception
 java.io.IOException:org.apache.parquet.crypto.KeyAccessDeniedException: Footer key:
 access denied
```

In-transit encryption in HiveServer2

Starting with Amazon EMR release 6.9.0, HiveServer2 (HS2) is TLS/SSL-enabled as part of In-transit encryption in HiveServer2 security configuration. This affects how you connect to HS2 running on an Amazon EMR cluster with in-transit encryption enabled. To connect to HS2, you must modify the `TRUSTSTORE_PATH` and `TRUSTSTORE_PASSWORD` parameter values in the JDBC URL. The following URL is an example of a JDBC connection for HS2 with the required parameters:

```
jdbc:hive2://HOST_NAME:10000/
default;ssl=true;sslTrustStore=TRUSTSTORE_PATH;trustStorePassword=TRUSTSTORE_PASSWORD
```

Use the appropriate instuctions for on-cluster or off-cluster HiveServer2 encryption below.

On-cluster HS2 access

If you are accessing HiveServer2 using the Beeline client after you SSH to the primary node, then reference `/etc/hadoop/conf/ssl-server.xml` to find the `TRUSTSTORE_PATH` and `TRUSTSTORE_PASSWORD` parameter values using configuration `ssl.server.truststore.location` and `ssl.server.truststore.password`.

The following example commands can help you retrieve these configurations:

```
TRUSTSTORE_PATH=$(sed -n '/ssl.server.truststore.location/,+2p' /etc/hadoop/conf/
ssl-server.xml | awk -F "[><]" '/value/{print $3}')
TRUSTSTORE_PASSWORD=$(sed -n '/ssl.server.truststore.password/,+2p' /etc/hadoop/
conf/ssl-server.xml | awk -F "[><]" '/value/{print $3}')
```

Off-cluster HS2 access

If you are accessing HiveServer2 from a client outside the Amazon EMR cluster. you can use one of the following approaches to get the TRUSTSTORE_PATH and TRUSTSTORE_PASSWORD:

- Convert the PEM file that was created during security configuration to a JKS file and use the same in the JDBC connection URL. For example, with openssl and keytool, use the following commands:

```
openssl pkcs12 -export -in trustedCertificates.pem -inkey privateKey.pem -out
 trustedCertificates.p12 -name "certificate"
keytool -importkeystore -srckeystore trustedCertificates.p12 -srcstoretype pkcs12
 -destkeystore trustedCertificates.jks
```

- Alternatively, reference /etc/hadoop/conf/ssl-server.xml to find the TRUSTSTORE_PATH and TRUSTSTORE_PASSWORD parameter values using configuration ssl.server.truststore.location and ssl.server.truststore.password. Download the truststore file to the client machine and use the path on the client machine as the TRUSTSTORE_PATH.

  For more information on accessing applications from a client outside of the Amazon EMR cluster, see Use the Hive JDBC driver.

Hive release history

The following table lists the version of Hive included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in Amazon EMR 7.x release versions, Amazon EMR 6.x release versions, or Amazon EMR 5.x release versions.

Hive version information

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-7.2.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server
emr-5.36.2	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server,

Amazon EMR Release label	Hive Version	Components installed with Hive
		hive-server2, hudi, mariadb-server, tez-on-yarn
emr-7.1.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-7.0.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-6.15.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-6.14.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-6.13.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-6.12.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-6.11.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-6.11.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-6.10.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-6.10.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-6.9.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.9.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-6.8.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.8.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-6.7.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-5.36.1	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.36.0	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-6.6.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.35.0	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-6.5.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-6.4.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.3.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-6.3.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.2.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-6.2.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.1.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-6.1.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.0.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-6.0.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-5.34.0	2.3.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.33.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.33.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.32.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.32.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.31.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.31.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.30.2	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.30.1	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.30.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.29.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.28.1	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mysql-server, tez-on-yarn
emr-5.28.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.27.1	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.27.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.26.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.25.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.24.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.24.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.23.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.23.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.22.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.21.2	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.21.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.21.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.20.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.20.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.19.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.19.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.18.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.18.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.17.2	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.17.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.17.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.16.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.16.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.15.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.15.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.14.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.14.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.14.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.13.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.13.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.12.3	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.12.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.12.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.12.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.11.4	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.11.3	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.11.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.11.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.11.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.10.1	2.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.10.0	2.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.9.1	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.9.0	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.8.3	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.8.2	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.8.1	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.8.0	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.7.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.7.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.6.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.6.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.5.4	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.5.3	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.5.2	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.5.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.5.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.4.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.4.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.3.2	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.3.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.3.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.2.3	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.2.2	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.2.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.2.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.1.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.1.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.0.3	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.0.2	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.0.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-5.0.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-4.9.6	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-4.9.5	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.9.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-4.9.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.9.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-4.9.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.8.5	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-4.8.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.8.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-4.8.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.8.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-4.8.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.7.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-4.7.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.7.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-4.7.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.7.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-4.6.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server
emr-4.6.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-4.5.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server
emr-4.4.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-4.3.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server
emr-4.2.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

Amazon EMR Release label	Hive Version	Components installed with Hive
emr-4.1.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server
emr-4.0.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

Hive release notes by version

Topics

Amazon EMR 7.1.0 - Hive release notes

Amazon EMR 7.1.0 - Hive changes

Type	Description
Bug Fix	HIVE-24381 – Compressed text input returns 0 rows if skip header/footer is included.
Bug Fix	HIVE-24190 – LLAP: ShuffleHandler might return DISK_ERROR_EXCEPTION according to TEZ-4233.
Bug Fix	HIVE-23073 – Shade Netty.
Bug Fix	HIVE-23073 – Shade Netty and upgrade to netty 4.1.48.Final.
Bug Fix	HIVE-23148 – Llap external client flow is broken due to netty shading.
Bug Fix	HIVE-25180 – Upgrades Netty.
Bug Fix	HIVE-24524 – LLAP ShuffleHandler: upgrade to Netty4 and remove Netty3 dependency from hive where it's possible.
Bug Fix	HIVE-28000 – Hive QL: the"not in" clause gives incorrect results when type coercion cannot take place.

Type	Description
Bug Fix	HIVE-27993 – Netty4 ShuffleHandler should use 1 boss thread.
Upgrade	Upgrades Netty to 4.1.100.Final
Upgrade	Upgrades Jetty to 9.4.53.v20231009
Upgrade	Upgrades Zookeeper to 3.9.1

Amazon EMR 7.1.0 - Hive changes

- Amazon EMR 7.1 upgrades Hive to Netty 4.1.100.Final to solve the security vulnerabilities in Netty3. Since hive-druid-handler has a dependency on netty3, Hive doesn't have the `hive-druid-handler` JAR in Hive's classpath in Amazon EMR 7.1. An upcoming Amazon EMR release will include it in Hive's classpath once the Druid handler supports 4.1.100.Final or later versions of Netty. Reach out to AWS support if you need the `hive-druid-handler` JAR in Amazon EMR releases 7.1 or higher.

Amazon EMR 6.15.0 - Hive release notes

Amazon EMR 6.15.0 - Hive changes

Type	Description
Feature	Support for TEZ-4397 – For Tez asynchronous split opening, Hive now supports the Tez configs described in Tez asynchronous split opening.
Bug fix	HIVE-25400 – Move the offset updating in `BytesColumnVector` to `setValPreallocated`.
Bug fix	HIVE-25190 – Fix many small allocations in `BytesColumnVector`.

Type	Description
Upgrade	Upgrade Apache Hadoop to 3.3.6.
Upgrade	HIVE-26684 – Upgrade `maven-shade-plugin` to 3.4.1.
Improvement	To reduce Amazon EMR cluster startup time, remove 15 seconds of sleep time from the HCatalog startup script.

Amazon EMR 6.14.0 - Hive release notes

Amazon EMR 6.14.0 - Hive changes

Type	Description
Improvement	HIVE-26762: Remove operand pruning in `HiveFilterSetOpTransposeRule`
Bug fix	HIVE-27582: Do not cache HBase table input format in FetchOperator
Bug fix	HIVE-26452: NPE when converting JOIN to MAPJOIN and JOIN column referenced more than once
Bug fix	HIVE-26416: `AcidUtils.isRawFormatFile()` throws `InvalidProtocolBufferException` for non-ORC file
Bug fix	HIVE-26105: **Show columns** shows extra values if column **Comments** contains specific Chinese character

Type	Description
Bug fix	HIVE-25864: Hive query optimisation creates wrong plan for predicate pushdown with windowing function
Bug fix	HIVE-25224: Multiple INSERT statements involving tables with different `bucketing_versions` results in error
Bug fix	HIVE-24151: `MultiDelimitSerDe` shifts data if strings contain non-ASCII characters
Bug fix	HIVE-23606: (LLAP) Delay In `DirectByteBuffer` cleanup for `EncodedReaderImpl`
Bug fix	HIVE-22165: Synchronisation introduced by HIVE-14296 on `SessionManager.closeSession` causes high latency in a busy Hive server
Bug fix	HIVE-21304: Make bucketing version usage more robust

Amazon EMR 6.13.0 - Hive release notes

Amazon EMR 6.13.0 - Hive changes

Type	Description
Improvement	Upgrade Python Scripts to Support Python3
Improvement	HIVE-27097: Improve the retry strategy for MetaStore client and server

Type	Description
Bug Fix	HIVE-21778: CBO: “Struct is not null” gets evaluated as nullable always causing filter miss in the query
Bug Fix	HIVE-21009: Adding ability for user to set bind user
Bug Fix	HIVE-22661: Compaction fails on non bucketed table with data loaded inpath
Bug Fix	HIVE-19718: Adding partitions in bulk also fetches table for each partition
Bug Fix	HIVE-22173: Query with multiple lateral views hangs during compilation
Bug Fix	HIVE-27088: Incorrect results when inner and outer joins with post join filters are merged
Bug Fix	HIVE-21935: Hive Vectorization : degraded performance with vectorize UDF
Bug Fix	HIVE-25299: Casting timestamp to numeric data types is incorrect for non-UTC timezones
Bug Fix	HIVE-24626: LLAP: reader threads could be starvated if all IO elevator threads are busy to enqueue to another readers with full queue
Bug Fix	HIVE-27029: hive query fails with Filesystem closed error, Rework done for HIVE-26352
Bug Fix	HIVE-26352: Tez queue access check fails with GSS Exception on Compaction
Bug Fix	HIVE-24590: Operation logging still leaks log4j appenders

Type	Description
Bug Fix	HIVE-24552: Possible HMS connections leak or accumulation in loadDynamicPartitions
Bug Fix	HIVE-27069: Incorrect results with bucket map join
Bug Fix	HIVE-27344: Add a null check in RecordRea derImpl#close
Bug Fix	HIVE-27439: Support space in Decimal
Bug Fix	HIVE-27267: Incorrect results when doing bucket map join on decimal bucketed column with subquery
Bug Fix	HIVE-21986: HiveServer Web UI: Setting the Strict-Transport-Security in default response header
Bug Fix	HIVE-22148: S3A delegation tokens are not added in the job config of the Compactor.
Bug Fix	HIVE-22622: Hive allows to create a struct with duplicate attribute names
Bug Fix	HIVE-22008: LIKE Operator should match multi-line input
Bug Fix	HIVE-23144: LLAP: Let QueryTracker cleanup on serviceStop
Bug Fix	HIVE-22391: NPE while checking Hive query results cache
Bug Fix	HIVE-23305: NullPointerException in LlapTaskSchedulerService addNode due to race condition

Type	Description
Bug Fix	HIVE-22178: Parquet FilterPredicate throws CastException after SchemaEvolution
Bug Fix	HIVE-21517: Fix AggregateStatsCache
Bug Fix	HIVE-21825: Improve client error msg when Active/Passive HA is enabled
Bug Fix	HIVE-23389: FilterMergeRule can lead to AssertionError
Bug Fix	HIVE-22767: Beeline doesn't parse semicolons in comments properly
Bug Fix	HIVE-22996: BasicStats parsing should check proactively for null or empty string
Bug Fix	HIVE-22808: HiveRelFieldTrimmer does not handle HiveTableFunctionScan
Bug Fix	HIVE-22437: LLAP Metadata cache NPE on locking metadata.
Bug Fix	HIVE-22606: AvroSerde logs avro.sche ma.literal under INFO level
Bug Fix	HIVE-22713: Constant propagation shouldn't be done for Join-Fil(*)-RS structure
Bug Fix	HIVE-21624: LLAP: Cpu metrics at thread level is broken
Bug Fix	HIVE-22815: reduce the unnecessary file system object creation in MROutput
Bug Fix	HIVE-23060: Query failing with error "Grouping sets expression is not in GROUP BY key. Error encountered near token"

Type	Description
Bug Fix	HIVE-22236: Fail to create View selecting View containing NOT IN subquery
Bug Fix	HIVE-19886: Logs may be directed to 2 files if —hiveconf hive.log.file is used
Bug Fix	HIVE-20620: manifest collisions when inserting into bucketed sorted MM tables with dynamic partitioning
Bug Fix	HIVE-14557: Nullpointer When both SkewJoin and Mapjoin Enabled
Bug Fix	HIVE-20471: issues getting the default database path
Bug Fix	HIVE-20598: Fix typos in HiveAlgorithmsUtil calculations
Bug Fix	HIVE-14737: Problem accessing /logs in a Kerberized Hive Server 2 Web UI
Bug Fix	HIVE-20733: GenericUDFOPEqualNS may not use = in plan descriptions
Bug Fix	HIVE-20848: After setting UpdateInp utAccessTimeHook query fail with Table Not Found.
Bug Fix	HIVE-18929: The method humanReadableInt in HiveStringUtils.java has a race condition.
Bug Fix	HIVE-20841: LLAP: Make dynamic ports configurable
Bug Fix	HIVE-20930: VectorCoalesce in FILTER mode doesn't take effect

Type	Description
Bug Fix	HIVE-21007: Semi join + Union can lead to wrong plans
Bug Fix	HIVE-21074: Hive bucketed table query pruning does not work for IS NOT NULL condition
Bug Fix	HIVE-21223: CachedStore returns null partition when partition does not exist
Bug Fix	HIVE-19625: Potential NPE and hiding actual exception in Hive#copyFiles
Bug Fix	HIVE-17020: Aggressive RS dedup can incorrectly remove OP tree branch
Bug Fix	HIVE-20168: ReduceSinkOperator Logging Hidden
Bug Fix	HIVE-20879: Using null in a projection expression leads to CastException
Bug Fix	HIVE-20888: TxnHandler: sort() called on immutable lists
Bug Fix	HIVE-19948: HiveCli is not splitting the command by semicolon properly if quotes are inside the string
Bug Fix	HIVE-20621: GetOperationStatus called in resultset.next causing incremental slowness
Bug Fix	HIVE-20854: Sensible Defaults: Hive's Zookeeper heartbeat interval is 20 minutes, change to 2

Type	Description
Bug Fix	HIVE-20330: HCatLoader cannot handle multiple InputJobInfo objects for a job with multiple inputs
Bug Fix	HIVE-20787: MapJoinBytesTableContainer dummyRow case doesn't handle reuse
Bug Fix	HIVE-20331: Query with union all, lateral view and Join fails with "cannot find parent in the child operator"
Bug Fix	HIVE-19968: UDF exception is not throw out
Bug Fix	HIVE-20410: aborted Insert Overwrite on transactional table causes "Not enough history available for…" error
Bug Fix	HIVE-20059: Hive streaming should try shade prefix unconditionally on exception
Bug Fix	HIVE-19424: NPE In MetaDataFormatters
Bug Fix	HIVE-20355: Clean up parameter of HiveConnection.setSchema
Bug Fix	HIVE-20858: Serializer is not correctly initialized with configuration in Utilities.createEmptyBuckets
Bug Fix	HIVE-20424: schematool shall not pollute beeline history
Bug Fix	HIVE-20338: LLAP: Force synthetic file-id for filesystems which have HDFS protocol impls with POSIX mutation semantics

Type	Description
Bug Fix	HIVE-11708: Logical operators raises ClassCast Exceptions with NULL
Bug Fix	HIVE-21082: In HPL/SQL, declare statement does not support variable of type character
Bug Fix	HIVE-16690: Configure Tez cartesian product edge based on LLAP cluster size
Bug Fix	HIVE-21296: Dropping varchar partition throw exception
Bug Fix	HIVE-14516: OrcInputFormat.SplitGenerat or.callInternal
Bug Fix	HIVE-20981: streaming/AbstractRecordWriter leaks HeapMemoryMonitor
Bug Fix	HIVE-20043: HiveServer2: SessionState has a static sync block around an AtomicBoolean
Bug Fix	HIVE-20191: PreCommit patch application doesn't fail if patch is empty
Bug Fix	HIVE-20400: create table should always use a fully qualified path to avoid potential FS ambiguity
Bug Fix	Add null check for skewedInfo before accessing skewed columns

Amazon EMR 6.12.0 - Hive release notes

Amazon EMR 6.12.0 - Hive changes

Type	Description
Improvement	Added Support For JDK 11 and JDK 17 Runtime
Improvement	Add support to query case sensitive and reserved keyword column names when using S3 Select. To use it, define table property in format "s3select.column.mapping" = "*column1*:*fieldName1* , *column2*:*fieldName2* ,..."
Improvement	HIVE-23133: Numeric operations can have different result across hardware archs
Improvement	HIVE-27145: Use StrictMath for remaining Math functions as followup of HIVE-23133
Bug Fix	Fix wildcard incompatibility in get_partitions_by_filter and get_num_partitions_by_filter HMS APIs caused by porting HIVE-22900 in EMR Hive 6.4.0
Bug Fix	HIVE-26736: Authorization failure for nested Views having WITH clause
Bug Fix	HIVE-22416: MR-related operation logs missing when parallel execution is enabled
Bug Fix	HIVE-19653: Incorrect predicate pushdown for groupby with grouping sets
Bug Fix	HIVE-22094: queries failing with ClassCast Exception: hive.ql.exec.vector.Decimal

Type	Description
	ColumnVector cannot be cast to hive.ql.e xec.vector.Decimal64ColumnVector
Bug Fix	HIVE-26340: Vectorized PTF operator fails if query has upper case window function
Bug Fix	HIVE-26184: COLLECT_SET with GROUP BY is very slow when some keys are highly skewed
Bug Fix	HIVE-26373: ClassCastException when reading timestamps from HBase table with Avro data
Bug Fix	HIVE-26388: ClassCastException when there is non string type column in source table of CTAS query Upgrade HIVE-26172: Hive - Upgrade Ant to 1.10.11 due to CVE-2021-36373 and CVE-2021-36374
Bug Fix	HIVE-26114: Fix jdbc connection hiveserve r2 using dfs command with prefix space will cause exceptio
Bug Fix	HIVE-26396: The trunc function has a problem with precision interception and the result has many 0
Bug Fix	HIVE-26446: HiveProtoLoggingHook fails to populate TablesWritten field for partitioned tables.
Bug Fix	HIVE-26639: ConstantVectorExpression and ExplainTask shouldn't rely on default charset
Bug Fix	HIVE-22670: ArrayIndexOutOfBoundsExcept ion when vectorized reader is used for reading a parquet file

Type	Description
Bug Fix	HIVE-23607: Permission Issue: Create view on another view succeeds but alter view fails
Bug Fix	HIVE-25498: Query with more than 31 count distinct functions returns wrong result
Bug Fix	HIVE-25780: DistinctExpansion creates more than 64 grouping sets II
Bug Fix	HIVE-23868: Windowing function spec: support 0 preceeding/following
Bug Fix	HIVE-24539: OrcInputFormat schema generation should respect column delimiter
Bug Fix	HIVE-23476: LLAP: Preallocate arenas for mmap case as well
Bug Fix	HIVE-25806: Possible leak in LlapCache AwareFs - Parquet, LLAP IO
Bug Fix	HIVE-23498: Disable HTTP Trace method on ThriftHttpCliService
Bug Fix	HIVE-25729: ThriftUnionObjectInspector should be notified when fully inited
Bug Fix	HIVE-23846: Avoid unnecessary serialization and deserialization of bitvectors
Bug Fix	HIVE-24233: except subquery throws nullpoint er with cbo disabled
Bug Fix	HIVE-24276: HiveServer2 loggerconf jsp Cross-Site Scripting (XSS) Vulnerability
Bug Fix	HIVE-25721: Outer join result is wrong

Type	Description
Bug Fix	HIVE-25223: Select with limit returns no rows on non native table
Bug Fix	HIVE-25794: CombineHiveRecordReader: log statements in a loop leads to memory pressure
Bug Fix	HIVE-23602: Use Java Concurrent Package for Operation Handle Set
Bug Fix	HIVE-24045: No logging related to when default database is created
Bug Fix	HIVE-24305: avro decimal schema is not properly populating scale/precision if value is enclosed in quote
Bug Fix	HIVE-25844: Exception deserialization error-s may cause beeline to terminate immediately
Bug Fix	HIVE-25040: Drop database cascade cannot remove persistent functions
Bug Fix	HIVE-23501: AOOB in VectorDeserializeR ow when complex types are converted to primitive types
Bug Fix	HIVE-23704: Thrift HTTP Server Does Not Handle Auth Handle Correctly
Bug Fix	HIVE-23529: CTAS is broken for uniontype when row_deserialize
Bug Fix	HIVE-24144: getIdentifierQuoteString in HiveDatabaseMetaData returns incorrect value

Type	Description
Bug Fix	HIVE-23850: Allow PPD when subject is not a column with grouping sets present
Bug Fix	HIVE-24036: Kryo Exception while serializing plan for getSplits UDF call
Bug Fix	HIVE-25919: ClassCastException when pushing boolean column predicate in HBaseStorageHandler
Bug Fix	HIVE-25261: RetryingHMSHandler should wrap the MetaException with short description of the target
Bug Fix	HIVE-24792: Potential thread leak in Operation
Bug Fix	HIVE-23409: If TezSession application reopen fails for Timeline service down, default TezSession from SessionPool is closed after a retry
Bug Fix	HIVE-23615: Do not deference null pointers in Beeline Commands Class
Bug Fix	HIVE-24849: Create external table socket timeout when location has large number of files (affects 3.1.2)
Bug Fix	HIVE-24193: Select query on renamed hive acid table does not produce any output
Bug Fix	HIVE-25209: SELECT query with SUM function producing unexpected result

Type	Description
Bug Fix	HIVE-23666: checkHashModeEfficiency is skipped when a groupby operator doesn't have a grouping set
Bug Fix	HIVE-23873: Querying Hive JDBCStora geHandler table fails with NPE when CBO is off
Bug Fix	HIVE-24149: HiveStreamingConnection doesn't close HMS connection
Bug Fix	HIVE-25561: Killed task should not commit file. (affects 2.x and 3.x versions)
Bug Fix	HIVE-25683: Close reader in AcidUtils .isRawFormatFile
Bug Fix	HIVE-24294: TezSessionPool sessions can throw AssertionError
Bug Fix	HIVE-24182: Ranger authorization issue with permanent UDFs
Bug Fix	HIVE-22805: Vectorization with conditional array or map is not implemented and throws an error
Bug Fix	HIVE-22828: Decimal64: NVL and CASE statements implicitly convert decimal64 to 128
Bug Fix	HIVE-21398: Columns which has estimated statistics should not be considered as unique keys
Bug Fix	HIVE-22490: Adding jars with special character s in their path throws error

Type	Description
Bug Fix	HIVE-22700: Compactions may leak memory when unauthorized
Bug Fix	HIVE-22053: Function name is not normalized when creating function
Bug Fix	HIVE-22595: Dynamic partition inserts fail on Avro table table with external schema
Bug Fix	HIVE-21795: Rollup summary row might be missing when a mapjoin is happening on a partitioned table
Bug Fix	HIVE-22987: ClassCastException in VectorCoa lesce when DataTypePhysicalVariation is null
Bug Fix	HIVE-22814: ArrayIndexOutOfBound in the vectorization getDataTypePhysicalVariation
Bug Fix	HIVE-22523: The error handler in LlapRecor dReader might block if its queue is full
Bug Fix	HIVE-21796: ArrayWritableObjectInspecto r.equals can take O(2^nesting_depth) time
Bug Fix	HIVE-22929: Performance: quoted identifie r parsing uses throwaway Regex via String.re placeAll()
Bug Fix	HIVE-21641: Llap external client returns decimal columns in different precision/scale as compared to beeline
Bug Fix	HIVE-22207: Tez: SplitGenerator throws NumberFormatException when "dfs.block.size" on cluster is "128m"

Type	Description
Bug Fix	HIVE-22114: Insert query for partitioned insert only table failing when all buckets are empty
Bug Fix	HIVE-22165: Synchronisation introduced by HIVE-14296 on SessionManager.closeSession causes high latency in a busy hive server
Bug Fix	HIVE-22744: TezTask for the vertex with more than one outedge should have proportional sort memory
Bug Fix	HIVE-22072: Altering table to make a column change does not update constraints references
Bug Fix	HIVE-22075: Fix the max-reducers=1 regressio n from HIVE-14200
Bug Fix	HIVE-22527: Hive on Tez : Job of merging small files will be submitted into another queue (default queue)
Bug Fix	HIVE-22816: QueryCache: Queries using views can have them cached after CTE expansion
Bug Fix	HIVE-22733: After disable operation log property in hive, still HS2 saving the operation log
Bug Fix	HIVE-22699: Mask UDFs should mask numeric value 0
Bug Fix	HIVE-23356: Hash aggregation is always disabled while processing querys with grouping sets expressions.

Type	Description
Bug Fix	HIVE-21568: HiveRelOptUtil.isRowFilteringPlan should skip Project
Bug Fix	HIVE-21760: Sharedwork optimization should be bypassed for SMB joins
Bug Fix	HIVE-22712: ReExec Driver execute submit the query in default queue irrespective of user defined queue
Bug Fix	HIVE-21397: BloomFilter for hive Managed [ACID] table does not work as expected
Bug Fix	HIVE-23011: Shared work optimizer should check residual predicates when comparing joins
Bug Fix	HIVE-21412: PostExecOrcFileDump doesn't work with ACID tables
Bug Fix	HIVE-22201: ConvertJoinMapJoin#checkShuffleSizeForLargeTable throws ArrayIndexOutOfBoundsException if no big table is selected
Bug Fix	HIVE-21971: HS2 leaks classloader due to \`ReflectionUtils::CONSTRUCTOR_CACHE\` with temporary functions + GenericUDF
Bug Fix	HIVE-21368: Vectorization: Unnecessary Decimal64 ->HiveDecimal conversion
Bug Fix	HIVE-25416: Hive metastore memory leak because datanucleus-api-jdo bug
Bug Fix	HIVE-22219: Bringing a node manager down blocks restart of LLAP service

Type	Description
Bug Fix	HIVE-21793: CBO retrieves column stats even if hive.stats.fetch.column.stats is set to false
Bug Fix	HIVE-22163: CBO: Enabling CBO turns on stats estimation, even when the estimation is disabled
Bug Fix	HIVE-18735: Create table like loses transacti onal attribute
Bug Fix	HIVE-22433: Hive JDBC Storage Handler: Incorrect results fetched from BOOLEAN and TIMESTAMP DataType From JDBC Data Source
Bug Fix	HIVE-19430: ObjectStore.cleanNotificati onEvents OutOfMemory on large number of pending events
Bug Fix	HIVE-20785: Wrong key name in the JDBC DatabaseMetaData.getPrimaryKeys method
Bug Fix	HIVE-16116: Beeline throws NPE when beeline.hiveconfvariables={} in beeline.p roperties
Bug Fix	HIVE-20066: hive.load.data.owner is compared to full principal
Bug Fix	HIVE-20489: Explain plan of query hangs
Bug Fix	HIVE-21033: Forgetting to close operation cuts off any more HiveServer2 output
Bug Fix	HIVE-19888: Misleading "METASTORE_FILTER_HOOK will be ignored" warning from SessionState

Type	Description
Bug Fix	HIVE-20303: INSERT OVERWRITE TABLE db.table PARTITION (...) IF NOT EXISTS throws InvalidTableException
Bug Fix	HIVE-16144: CompactionInfo doesn't have equals/hashCode but used in Set
Bug Fix	HIVE-20818: Views created with a WHERE subquery will regard views referenced in the subquery as direct input
Bug Fix	HIVE-21005: LLAP: Reading more stripes per-split leaks ZlibCodecs
Bug Fix	HIVE-20771: LazyBinarySerDe fails on empty structs.
Bug Fix	HIVE-18852: Misleading error message in alter table validation
Bug Fix	HIVE-21124: HPL/SQL does not support the CREATE TABLE LIKE statement
Bug Fix	HIVE-20935: Upload of llap package tarball fails in EC2 causing LLAP service start failure
Bug Fix	HIVE-20409: Hive ACID: Update/delete/merge does not clean hdfs staging directory
Bug Fix	HIVE-20570: Union ALL with hive.optimize.union.remove=true has incorrect plan
Bug Fix	HIVE-20421: Illegal character entity '\b' in hive-default.xml.template
Bug Fix	HIVE-19133: HS2 WebUI phase-wise performance metrics not showing correctly

Type	Description
Bug Fix	HIVE-18977: Listing partitions returns different results with JDO and direct SQL
Bug Fix	HIVE-20034: Roll back MetaStore exception handling changes for backward compatibility
Bug Fix	HIVE-20672: Logging thread in LlapTaskSchedulerService should report every fixed interval
Bug Fix	HIVE-12812: Enable mapred.input.dir.recursive by default to support union with aggregate function
Bug Fix	HIVE-20147: Hive streaming ingest is contented on synchronized logging
Bug Fix	HIVE-19203: Thread-Safety Issue in HiveMetaStore
Bug Fix	HIVE-20091: Tez: Add security credentials for FileSinkOperator output
Bug Fix	HIVE-16906: Hive ATSHook should check for yarn.timeline-service.enabled before connecting to ATS
Bug Fix	HIVE-20714: SHOW tblproperties for a single property returns the value in the name column
Bug Fix	HIVE-24730: Shims classes override values from hive-site.xml and tez-site.xml silently
Bug Fix	HIVE-22055: select count gives incorrect result after loading data from text file

Amazon EMR 6.11.0 - Hive release notes

Amazon EMR 6.11.0 - Hive changes

Type	Description
Improvement	Added support for multithreaded dropping of partitions to improve the performance of dropping of partitions
Improvement	Support reading encoded Hive query files
Improvement	Enabled Tez Shuffle Handler by default for Hive on Tez jobs
Bug	Added an option to enable deterministic distribution of keys to reducers to fix incorrect result when hive.groupby.skewindata is enabled (reported in HIVE-20220)
Bug	Fixed stats computation failure when default partition name is configured
Bug	Respect any custom SSL classification parameters passed when SSL is configured out of the box for HiveServer2 in a cluster with in-transit encryption enabled
Backport	HIVE-23617: Fixed storage-api FindBug issues
Backport	HIVE-26408: Vectorization: Fix deallocation of scratch columns, don't reuse a child ConstantVectorExpression as an output
Backport	HIVE-23614: Always pass HiveConfig to removeTempOrDuplicateFiles
Backport	HIVE-23354: Remove file size sanity checking from compareTempOrDuplicateFiles

Type	Description
Backport	HIVE-20344: Fixed PrivilegeSynchronizer for SBA throwing AccessControlException. Also introduced property hive.privilege.syn chronizer to disable privilege synchronizer
Backport	HIVE-15826: Support configuring 'serializ ation.encoding' for all SerDes
Backport	HIVE-18284: Fix NPE when inserting data with 'distribute by' clause with dynpart sort optimization
Backport	HIVE-24930: Operator.setDone() short-cir cuit from child op is not used in vectorized codepath (if childSize == 1)
Backport	HIVE-24523: Vectorized read path for LazySimpleSerde does not honor the SERDEPROPERTIES for timestamp
Backport	HIVE-23265: Duplicate rowsets are returned with Limit and Offset set
Backport	HIVE-21492: VectorizedParquetRecordRead er can't to read parquet file generated using thrift/custom tool
Backport	HIVE-22540: Vectorization: Decimal64 columns don't work with VectorizedBatchUti l.makeLikeColumnVector()
Backport	HIVE-22588: Flush the remaining rows for the rest of the grouping sets when switching the vector groupby mode
Backport	HIVE-22551: BytesColumnVector initBuffer should clean vector and length consistently

Type	Description
Backport	HIVE-22448: CBO: Expand the multiple count distinct with a group-by key
Backport	HIVE-22248: Fix statistics persisting issues
Backport	HIVE-22210: Vectorization may reuse computation output columns involved in filtering
Backport	HIVE-21531: Vectorization: all NULL hashcodes are not computed using Murmur3
Backport	HIVE-20419: Vectorization: Prevent mutation of VectorPartitionDesc after being used in a hashmap key
Backport	HIVE-19388: ClassCastException during VectorMapJoinCommonOperator initialization
Backport	HIVE-21584: Java 11 preparation: system class loader is not URLClassLoader
Backport	HIVE-25107: Classpath logging should be on DEBUG level (#2271)
Backport	HIVE-22097: Incompatible java.util.ArrayList for java 11
Backport	HIVE-23938: LLAP: JDK11 - some GC log file rotation related jvm arguments cannot be used anymore
Backport	HIVE-26226: Exclude jdk.tools dep from hive-metastore in upgrade-acid
Backport	HIVE-17879: Upgrade Datanucleus Maven Plugin

Type	Description
Backport	HIVE-27004: DateTimeFormatterBuilder#appendZoneText cannot parse 'UTC+' in Java versions higher than 8
Backport	HIVE-16812: VectorizedOrcAcidRowBatchReader doesn't filter delete events
Backport	HIVE-17917: VectorizedOrcAcidRowBatchReader.computeOffsetAndBucket optimization
Backport	HIVE-19985: ACID: Skip decoding the ROW__ID sections for read-only queries
Backport	HIVE-20635: VectorizedOrcAcidRowBatchReader doesn't filter delete events for original files
Upgrade	Upgrade Javadoc to 3.3.1
Upgrade	Upgrade Javassist to 3.24.1-GA
Upgrade	Update apache-directory-server to 2.0.0-M14

New configurations

Name	Classification	Description
hive.metastore.fs.drop.partition.threads	hive-site	Number of core threads in the drop partition thread pool.
hive.metastore.fs.drop.partition.keepalive.time	hive-site	Time in seconds that an idle drop partition async thread (from the thread pool) will wait for a new task to arrive before terminating.

Name	Classification	Description
hive.metastore.fs.drop.partition.threadpool.max.queue.size	hive-site	Maximum queue size to be used in thread pool for dropping of partitions from file system.
hive.groupby.enable.deterministic.distribution	hive-site	Enable deterministic distribution of keys to reducers. It will pass a constant seed value while calling the rand function used for random partitioning.
hive.privilege.synchronizer	hive-site	Whether to synchronize privileges from external authorizer periodically in HiveServer2.
hive.cli.query.file.encoding	hive-site	File encoding for the all type of query files (query file, init query file, rc file etc) provided in the cli arguments.
hive.emr.tez.shuffle.enabled	hive-site	Hive on Tez jobs now use tez_shuffle by default instead of mapreduce_shuffle as the default Shuffle Handler.

Deprecated configurations

The following configuration properties are deprecated as a result of HIVE-23354 and are no longer supported with Amazon EMR releases 6.11.0 and higher.

Name	Default value
`hive.mapred.reduce.tasks.speculative.execution`	`false`
`tez.am.speculation.enabled`	`false`

Amazon EMR 6.10.0 - Hive release notes

Amazon EMR 6.10.0 - Hive changes

Type	Description
Feature	Enable AWS Lake Formation based access controls for Apache Hive queries **(write)** via IAM Passthrough (HiveCLI/Steps API).
Improvement	Disable config `hive.log.explain.output` by default to reduce log size
Backport	HIVE-26408: Vectorization: Fix deallocation of scratch columns, don't reuse a child ConstantVectorExpression as an output
Backport	HIVE-22269: Fix wrong reducers count in insert queries with dynamic partition due to missing of stats caused by HIVE-20703.
Backport	HIVE-22891: Skip PartitionDesc Extraction In CombineHiveRecord For Non-LLAP Execution Mode
Backport	HIVE-23804: Add default database for column stats specific tables in Hive metastore schema to make then backward compatible
Backport	HIVE-25277: Slow Hive partition deletion for Cloud object stores with expensive ListFiles

Type	Description
Backport	HIVE-19202: CBO failed due to NullPointerException in HiveAggregate.isBucketedInput()
Backport	HIVE-19048: Fix beeline Initscript errors are ignored
Backport	HIVE-21085: Materialized views registry starts non-external tez session
Backport	HIVE-21675: CREATE VIEW IF NOT EXISTS returns an error rather than "OK" if the view already exists. This is a regression from Hive 2.
Backport	HIVE-21646: Tez: Prevent TezTasks from escaping thread logging context
Backport	HIVE-22054: Avoid recursive listing to check if a directory is empty
Backport	HIVE-16587: NPE when inserting complex types with nested null values
Backport	HIVE-22647: Enable session pool by default
Backport	HIVE-13288: Confusing exception message in DagUtils.localizeResource
Backport	HIVE-23870: Optimise multiple text conversions in WritableHiveCharObjectInspector.getPrimitiveJavaObject / HiveCharWritable
Backport	HIVE-21498: Upgrade Thrift to 0.13.0
Backport	HIVE-24378: Leading and trailing spaces are not removed before decimal conversion

Type	Description
Backport	HIVE-21341: Sensible defaults : hive.serv er2.idle.operation.timeout and hive.serv er2.idle.session.timeout are too high
Backport	HIVE-22465: Add ssl conf in TezConfig urationFactory
Backport	HIVE-24710: Optimise PTF iteration for count(*) to reduce CPU and IO cost
Backport	HIVE-15406: Consider vectorizing the new 'trunc' function
Backport	HIVE-21541: Fix missing asf headers from HIVE-15406
Backport	HIVE-24808: Cache Parsed Dates
Backport	HIVE-24746: PTF: TimestampValueBoun daryScanner can be optimised during range computation
Backport	HIVE-25059: Alter event is converted to rename during replication
Backport	HIVE-25142: Rehashing in map join fast hash table causing corruption for large keys
Backport	HIVE-23756: Added more constraints to the package.jdo file
Backport	HIVE-25150: Tab characters are not removed before decimal conversion similar to space character which is fixed as part of HIVE-24378
Backport	HIVE-25093: date_format() UDF is returning output in UTC time zone only

Type	Description
Backport	HIVE-25268: date_format udf returns wrong results for dates prior to 1900 if the local timezone is other than UTC
Backport	HIVE-25338: AIOBE in conv UDF if input is empty
Backport	HIVE-22400: UDF minute with time returns NULL
Backport	HIVE-25058: PTF: TimestampValueBoun daryScanner can be optimised during range computation pt2 - isDistanceGreater
Backport	HIVE-25449: datediff() gives wrong output when run in a tez task with some non-UTC timezone
Backport	HIVE-23688: Vectorization: IndexArra yOutOfBoundsException For map type column which includes null value
Backport	HIVE-22247: HiveHFileOutputFormat throws FileNotFoundException when partition's task output empty
Backport	HIVE-25570: Hive should send full URL path for authorization for the command insert overwrite location
Backport	HIVE-22903: Vectorized row_number() resets the row number after one batch in case of constant expression in partition clause
Backport	HIVE-25549: Wrong results for window function with expression in PARTITION BY or ORDER BY clause

Type	Description
Backport	HIVE-25579: LOAD overwrite appends rather than overwriting
Backport	HIVE-25659: Metastore direct sql queries with IN/(NOT IN) should be split based on max parameters allowed by SQL DB
Backport	HIVE-20502: Fix NPE while running skewjoin_mapjoin10.q when column stats is used.
Backport	HIVE-25765: skip.header.line.count property skips rows of each block in FetchOperator when file size is larger
Bug	Fix NPE on insert in certain scenarios when `hive.stats.column.autogather` and `hive.groupby.skewindata` are both enabled
Bug	Fix NPE when `mapred.tasktracker.expiry.interval` value is not set

Amazon EMR 6.9.0 - Hive release notes

Amazon EMR 6.9.0 - Hive changes

Type	Description
Upgrade	Upgrade Jetty to 9.4.48.v20220622
Upgrade	Support for Hadoop 3.3.3
Feature	Amazon EMR Hive integration with Lake Formation for interactive workloads using GCSC API.

Type	Description
Feature	Amazon EMR Hive integration with Iceberg.
Improvement	Enable SSL in HiveServer2 when in-transit encryption is enabled using Amazon EMR security configurations.
Improvement	Enable Hive EMRFS Amazon S3 optimized committer by default. For more information see, Enabling Hive EMRFS S3 optimized committer.
Improvement	Add `HiveHBaseTableInputFormatV2` that inherits only mapred version of InputFormat to fix SPARK-34210. Set `hive.hbase.inputformat.v2` to `true` to use it.
Improvement	Wait for TezAM to launch in background with hive.cli.tez.session.async instead of terminating it and launching new immediately. Use `hive.emr.cli.tez.session.open.timeout` to set this timeout in seconds.
Improvement	Add option hive.conf.restricted.list.append to append comma-separated configs to existing restricted config list `hive.conf.restricted.list` .
Improvement	Clearer error message when Hive query fails because location is not defined for database.
Backport	HIVE-24484: Upgrade Hadoop to 3.3.1 And Tez to 0.10.2
Backport	HIVE-22398: Remove YARN queue management via ShimLoader.

Type	Description
Backport	HIVE-23190: LLAP: modify IndexCache to pass filesystem object to TezSpillRecord.
Backport	HIVE-22185: HADOOP-15832 will cause problems with tests using MiniYarn clusters.
Backport	HIVE-21670: Replacing `mockito-all` with `mockito-core` dependency.
Backport	HIVE-24542: Prepare Guava for Upgrades.
Backport	HIVE-23751: QTest: Override `#mkdirs()` method in `ProxyFileSystem` to align after HADOOP-16582.
Backport	HIVE-21603: Java 11 preparation: update powermock version.
Backport	HIVE-24083: `hcatalog` error in Hadoop 3.3.0: authentication type needed.
Backport	HIVE-24282: Show columns shouldn't sort output columns unless explicitly mentioned.
Backport	HIVE-20656: Sensible defaults: Map aggregati on memory configs are too aggressive.
Backport	HIVE-25443: Arrow SerDe cannot serialize/ deserialize complex data types when there are more than 1024 values
Backport	HIVE-19792: Upgrade orc to 1.5.2 and enable decimal_64 schema evolution tests.
Backport	HIVE-20437: Handle schema evolution from float, double, and decimal.

Type	Description
Backport	HIVE-21987: Hive is unable to read Parquet int32 annotated with decimal.
Backport	HIVE-20038: Update queries on non-bucketed and partitioned tables throws NPE.

Amazon EMR 6.9.0 - Hive known issues

- With Amazon EMR 6.6.0 through 6.9.x, INSERT queries with dynamic partition and an ORDER BY or SORT BY clause will always have two reducers. This issue is caused by OSS change HIVE-20703, which puts dynamic sort partition optimization under cost-based decision. If your workload doesn't require sorting of dynamic partitions, we recommend that you set the `hive.optimize.sort.dynamic.partition.threshold` property to `-1` to disable the new feature and get the correctly calculated number of reducers. This issue is fixed in OSS Hive as part of HIVE-22269 and is fixed in Amazon EMR 6.10.0.

Amazon EMR 6.8.0 - Hive release notes

Amazon EMR 6.8.0 - Hive changes

Type	Description
Improvement	Reduce file system calls in msck command. Performance improvements (~15-20x on 10k+ partitions)
Backport	HIVE-20678: HiveHBaseTableOutputFormat should implement HiveOutputFormat to ensure compatibility
Backport	HIVE-21040: msck does unnecessary file listing at last level of directory tree
Backport	HIVE-21460: Load data followed by a select * query results in incorrect results

Type	Description
Backport	HIVE-21660: Wrong result when union all and later view with explode is used
Backport	HIVE-22505: ClassCastException caused by wrong Vectorized operator selection
Backport	HIVE-22513: Constant propagation of casted column in filter ops can cause incorrect results
Backport	HIVE-23435: Full outer join result is missing rows
Backport	HIVE-24209: Incorrect search argument conversion for NOT BETWEEN operation when vectorization is enabled
Backport	HIVE-24934: VectorizedExpressions annotation is not needed in GenericUDFSQCountCheck
Backport	HIVE-25278: HiveProjectJoinTransposeRule may do invalid transformations with windowing expressions
Backport	HIVE-25505: Incorrect results with header.skip.header.line.count if first line is blank
Backport	HIVE-26080: Upgrade accumulo-core to 1.10.1
Backport	HIVE-26235: OR Condition on binary column is returning empty result
Bug	Fix multiple SLF4J bindings warning logs in stderr during launch
Bug	Fix SHOW TABLE EXTENDED query failing with Wrong FS error when partition and table are on different file systems.

Amazon EMR 6.8.0 - Hive known issues

- With Amazon EMR 6.6.0 through 6.9.x, INSERT queries with dynamic partition and an ORDER BY or SORT BY clause will always have two reducers. This issue is caused by OSS change HIVE-20703, which puts dynamic sort partition optimization under cost-based decision. If your workload doesn't require sorting of dynamic partitions, we recommend that you set the `hive.optimize.sort.dynamic.partition.threshold` property to `-1` to disable the new feature and get the correctly calculated number of reducers. This issue is fixed in OSS Hive as part of HIVE-22269 and is fixed in Amazon EMR 6.10.0.

Amazon EMR 6.7.0 - Hive release notes

Amazon EMR 6.7.0 - Hive changes

Type	Description
Feature	Amazon EMR Hive integration with LakeFormation.
Feature	Additional audit logging for Hive EMRFS Amazon S3 optimized committer. Hive config: `hive.blobstore.output-committer.logging` , default: `false`
Feature	Deleted target directory on insert overwrite with empty select result to an unpartitioned table/static partition to behave similarly to Hive 2.x. Hive config: `hive.emr.iow.clean.target.dir` , default: `false`
Bug	Fixed intermittent query failure when using Hive EMRFS Amazon S3 optimized committer with partition bucket sorting.
Upgrade	Hive version upgraded to 3.1.3. Refer to Apache Hive 3.1.3 release notes for more details.

Type	Description
Upgrade	Upgraded Parquet to 1.12.2.
Backport	HIVE-20065: Metastore should not rely on jackson 1.x
Backport	HIVE-20071: Migrate to jackson 2.x and prevent usage
Backport	HIVE-20607: TxnHandler should use PreparedStatement to execute direct SQL queries
Backport	HIVE-20740: Remove global lock in ObjectSto re.setConf method
Backport	HIVE-20961: Retire NVL implementation
Backport	HIVE-22059: hive-exec jar doesn't contain (fasterxml) jackson library
Backport	HIVE-22351: Fix incorrect threaded ObjectSto re usage in TestObjectStore
Backport	HIVE-23534: NPE in RetryingMetaStoreC lient#invoke when catching MetaException with no message
Backport	HIVE-24048: Harmonise Jackson components to version 2.10.latest - Hive
Backport	HIVE-24768: Use jackson-bom everywhere for version replacement
Backport	HIVE-24816: Upgrade jackson to 2.10.5.1 or 2.11.0+ due to CVE-2020-25649
Backport	HIVE-25971: Tez task shutdown getting delayed due to cached thread pool not closed

Type	Description
Backport	HIVE-26036: NPE caused by getMTable() in ObjectStore

Amazon EMR 6.7.0 - Hive known issues

- Queries with windowing functions on the same column as join may lead to invalid transformations as reported in HIVE-25278 and cause incorrect results or query failures. A workaround would be to disable CBO at the query level for such queries. The fix will be available in an Amazon EMR release following 6.7.0. For more information, contact AWS support.
- With Amazon EMR 6.6.0 through 6.9.x, INSERT queries with dynamic partition and an ORDER BY or SORT BY clause will always have two reducers. This issue is caused by OSS change HIVE-20703, which puts dynamic sort partition optimization under cost-based decision. If your workload doesn't require sorting of dynamic partitions, we recommend that you set the `hive.optimize.sort.dynamic.partition.threshold` property to `-1` to disable the new feature and get the correctly calculated number of reducers. This issue is fixed in OSS Hive as part of HIVE-22269 and is fixed in Amazon EMR 6.10.0.

Amazon EMR 6.6.0 - Hive release notes

Amazon EMR 6.6.0 - Hive changes

Type	Description
Upgrade	Upgrade Parquet to 1.12.1.
Upgrade	Upgrade jetty jars version to 9.4.43.v2 0210629
Bug	Fixed an issue that was causing Hive to be installed on all task/core nodes when LLAP was enabled on a Hive cluster.
Backport	HIVE-25942: Upgrade commons-io to 2.8.0 due to CVE-2021-29425

Type	Description
Backport	HIVE-25726: Upgrade velocity to 2.3 due to CVE-2020-13936
Backport	HIVE-25680: Authorize #get_table_meta HiveMetastore Server API to use any of the HiveMetastore Authorization model.
Backport	HIVE-25554: Upgrade arrow version to 0.15
Backport	HIVE-25242: Query performs extremely slow with vectorized.adaptor = chosen
Backport	HIVE-25085: MetaStore Clients no longer shared across sessions.
Backport	HIVE-24827: Hive aggregation query returns incorrect results for non text files.
Backport	HIVE-24683: Hadoop23Shims getFileId prone to NPE for non-existing paths
Backport	HIVE-24656: CBO fails for queries with is null on map and array types
Backport	HIVE-24556: Optimize DefaultGraphWalker for case with no grandchild
Backport	HIVE-24408: Upgrade Parquet to 1.11.1
Backport	HIVE-24391: Fix FIX TestOrcFile failures in branch-3.1
Backport	HIVE-24362: AST tree processing is suboptima l for tree with large number of nodes
Backport	HIVE-24316: Upgrade ORC from 1.5.6 to 1.5.8 in branch-3.1

Type	Description
Backport	HIVE-24307: Beeline with property-file and -e parameter is failing
Backport	HIVE-24245: Vectorized PTF with count and distinct over partition producing incorrect results.
Backport	HIVE-24224: Fix skipping header/footer for Hive on Tez on compressed file
Backport	HIVE-24157: Strict mode to fail on CAST timestamp ↔ numeric
Backport	HIVE-24113: NPE in GenericUDFToUnixTi meStamp
Backport	HIVE-23987: Upgrade arrow version to 0.11.0
Backport	HIVE-23972: Add external client ID to LLAP external client
Backport	HIVE-23806: Avoid clearing column stat states in all partition in case schema is extended. This improves runtime of alter table add columns statement.
Backport	HIVE-23779: BasicStatsTask Info is not getting printed in beeline console
Backport	HIVE-23306: RESET command does not work if there is a config set by System.getProperty
Backport	HIVE-23164: Server is not properly terminated because of non-daemon threads
Backport	HIVE-22967: Support hive.reloadable.au x.jars.path for Hive on Tez

Type	Description
Backport	HIVE-22934: Hive server interactive log counters to error stream
Backport	HIVE-22901: Variable substitution can lead to OOM on circular references
Backport	HIVE-22769: Incorrect query results and query failure during split generation for compressed text files
Backport	HIVE-22716: Reading to ByteBuffer is broken in ParquetFooterInputFromCache
Backport	HIVE-22648: Upgrade Parquet to 1.11.0
Backport	HIVE-22640: Decimal64ColumnVector: ClassCastException when partition column type is Decimal
Backport	HIVE-22621: unstable testcase: TestLlapS ignerImpl.testSigning
Backport	HIVE-22533: Fix possible LLAP daemon web UI vulnerabilities
Backport	HIVE-22532: PTFPPD may push limit incorrect ly through Rank/DenseRank function
Backport	HIVE-22514: HiveProtoLoggingHook might consume lots of memory
Backport	HIVE-22476: Hive datediff function provided inconsistent results when hive.fetch.task.co nversion is set to none

Type	Description
Backport	HIVE-22429: Migrated clustered tables using bucketing_version 1 on hive 3 uses bucketing _version 2 for inserts
Backport	HIVE-22412: StatsUtils throw NPE when explain
Backport	HIVE-22360: MultiDelimitSerDe returns wrong results in last column when the loaded file has more columns than those in table schema
Backport	HIVE-22332: Hive should ensure valid schema evolution settings since ORC-540
Backport	HIVE-22331: unix_timestamp without argument returns timestamp in millisecond instead of second
Backport	HIVE-22275: OperationManager.queryIdOpe ration does not properly clean up multiple queryIds
Backport	HIVE-22273: Access check is failed when a temporary directory is removed
Backport	HIVE-22270: Upgrade commons-io to 2.6
Backport	HIVE-22241: Implement UDF to interpret date/timestamp using its internal represent ation and Gregorian-Julian hybrid calendar
Backport	HIVE-22241: Implement UDF to interpret date/timestamp using its internal represent ation and Gregorian-Julian hybrid
Backport	HIVE-22232: NPE when hive.order.columna lignment is set to false

Type	Description
Backport	HIVE-22231: Hive query with big size via knox fails with Broken pipe Write failed
Backport	HIVE-22221: Llap external client - Need to reduce LlapBaseInputFormat#getSplits
Backport	HIVE-22208: Column name with reserved keyword is unescaped when query including join on table with mask column is re-written
Backport	HIVE-22197: Common Merge join throwing class cast exception.
Backport	HIVE-22170: from_unixtime and unix_time stamp should use user session time zone
Backport	HIVE-22169: Tez: SplitGenerator tries to look for plan files which won't exist for Tez
Backport	HIVE-22168: Remove very expensive logging from the llap cache hotpath
Backport	HIVE-22161: UDF: FunctionRegistry synchroni zes on org.apache.hadoop.hive.ql.udf.UDFType class
Backport	HIVE-22120: Fix wrong results/ArrayOutOf Bound exception in left outer map joins on specific boundary conditions
Backport	HIVE-22115: Prevent the creation of query routing appender if property is set to false
Backport	HIVE-22113: Prevent LLAP shutdown on AMReporter related RuntimeException

Type	Description
Backport	HIVE-22106: Remove cross-query synchronization for the partition-eval
Backport	HIVE-22099: Several date related UDFs can't handle Julian dates properly since HIVE-20007
Backport	HIVE-22037: HS2 should log when shutting down due to OOM
Backport	HIVE-21976: Offset should be null instead of zero in Calcite HiveSortLimit
Backport	HIVE-21924: Split text files even if header/footer exists
Backport	HIVE-21913: GenericUDTFGetSplits should handle usernames in the same way as LLAP
Backport	HIVE-21905: Generics improvement around the FetchOperator class
Backport	HIVE-21902: HiveServer2 UI: jetty response header needs X-Frame-Options
Backport	HIVE-21888: Set hive.parquet.timestamp.skip.conversion default to true
Backport	HIVE-21868: Vectorize CAST...FORMAT
Backport	HIVE-21864: LlapBaseInputFormat#closeAll
Backport	HIVE-21863: Improve Vectorizer type casting for WHEN expression
Backport	HIVE-21862: ORC ppd produces wrong result with timestamp

Type	Description
Backport	HIVE-21846: Create a thread in TezAM which periodically fetches LlapDaemon metrics
Backport	HIVE-21837: MapJoin is throwing exception when selected column is having completely null values
Backport	HIVE-21834: Avoid unnecessary calls to simplify filter conditions
Backport	HIVE-21832: New metrics to get the average queue/serving/response time
Backport	HIVE-21827: Multiple calls in SemanticA nalyzer do not go through getTableO bjectByName method
Backport	HIVE-21822: Expose LlapDaemon metrics through a new API method
Backport	HIVE-21818: CBO: Copying TableRelO ptHiveTable has metastore traffic
Backport	HIVE-21815: Stats in ORC file are parsed twice
Backport	HIVE-21805: HiveServer2: Use the fast ShutdownHookManager APIs
Backport	HIVE-21799: NullPointerException in DynamicPartitionPruningOptimization, when join key is on aggregation column
Backport	HIVE-21794: Add materialized view parameter s to sqlStdAuthSafeVarNameRegexes
Backport	HIVE-21768: JDBC: Strip the default union prefix for un-enclosed UNION queries

Type	Description
Backport	HIVE-21746: ArrayIndexOutOfBoundsException during dynamically partitioned hash join, with CBO disabled
Backport	HIVE-21717: Rename is failing for directory in move task.
Backport	HIVE-21685: Wrong simplification in query with multiple IN clauses
Backport	HIVE-21681: Describe formatted shows incorrect information for multiple primary keys
Backport	HIVE-21651: Move protobuf serde into hive-exec.
Backport	HIVE-21619: Print timestamp type without precision in SQL explain extended
Backport	HIVE-21592: OptimizedSql is not shown when the expression contains CONCAT
Backport	HIVE-21576: Introduce CAST...FORMAT and limited list of SQL:2016 datetime formats
Backport	HIVE-21573: Binary transport shall ignore principal if auth is set to delegationToken
Backport	HIVE-21550: TestObjectStore tests are flaky - A lock could not be obtained within the time requested
Backport	HIVE-21544: Constant propagation corrupts coalesce/case/when expressions during folding

Type	Description
Backport	HIVE-21539: GroupBy + where clause on same column results in incorrect query rewrite
Backport	HIVE-21538: Beeline: password source though the console reader did not pass to connection param
Backport	HIVE-21509: LLAP may cache corrupted column vectors and return wrong query result
Backport	HIVE-21499: should not remove the function from registry if create command failed with AlreadyExistsException
Backport	HIVE-21496: Automatic sizing of unordered buffer can overflow
Backport	HIVE-21468: Case sensitivity in identifier names for JDBC storage handler
Backport	HIVE-21467: Remove deprecated junit.framework.Assert imports
Backport	HIVE-21435: LlapBaseInputFormat should get task number from TASK_ATTEMPT_ID conf if present, while building SubmitWorkRequestProto
Backport	HIVE-21389: Hive distribution miss javax.ws.rs-api.jar after HIVE-21247
Backport	HIVE-21385: Allow disabling pushdown of non-splittable computation to JDBC sources
Backport	HIVE-21383: JDBC storage handler: Use catalog and schema to retrieve tables if specified

Type	Description
Backport	HIVE-21382: Group by keys reduction optimization - keys are not reduced in query23
Backport	HIVE-21362: Add an input format and serde to read from protobuf files.
Backport	HIVE-21340: CBO: Prune non-key columns feeding into a SemiJoin
Backport	HIVE-21332: Purge the non locked buffers instead of locked ones
Backport	HIVE-21329: Custom Tez runtime unordered output buffer size depending on operator pipeline
Backport	HIVE-21295: StorageHandler shall convert date to string using Hive convention
Backport	HIVE-21294: Vectorization: 1-reducer Shuffle can skip the object hash functions
Backport	HIVE-21255: Remove QueryConditionBuilder in JdbcStorageHandler
Backport	HIVE-21253: Support DB2 in JDBC StorageHa ndler
Backport	HIVE-21232: LLAP: Add a cache-miss friendly split affinity provider
Backport	HIVE-21214: MoveTask : Use attemptId instead of file size for deduplication of files compareTempOrDuplicateFiles
Backport	HIVE-21184: Add explain and explain formatted CBO plan with cost information

Type	Description
Backport	HIVE-21182: Skip setting up hive scratch dir during planning
Backport	HIVE-21171: Skip creating scratch dirs for tez if RPC is on
Backport	HIVE-21126: Allow session level queries in LlapBaseInputFormat#getSplit
Backport	HIVE-21107: Cannot find field" error during dynamically partitioned hash join
Backport	HIVE-21061: CTAS query fails with IllegalSt ateException for empty source
Backport	HIVE-21041: NPE, ParseException in getting schema from logical plan
Backport	HIVE-21013: JdbcStorageHandler fail to find partition column in Oracle
Backport	HIVE-21006: Extend SharedWorkOptimizer to remove semijoins when there is a reutilization opportunity
Backport	HIVE-20992: Split the config hive.meta store.dbaccess.ssl.properties into more meaningful configs
Backport	HIVE-20989: JDBC - The GetOperationStatus + log can block query progress via sleep
Backport	HIVE-20988: Wrong results for group by queries with primary key on multiple columns

Type	Description
Backport	HIVE-20985: If select operator inputs are temporary columns vectorization may reuse some of them as output
Backport	HIVE-20978: "hive.jdbc.*" should add to sqlStdAuthSafeVarNameRegexes
Backport	HIVE-20953: Remove a function from function registry when it can not be added to the metastore when creating it.
Backport	HIVE-20952: Cleaning VectorizationConte xt.java
Backport	HIVE-20951: LLAP: Set Xms to 50% always
Backport	HIVE-20949: Improve PKFK cardinality estimation in physical planning
Backport	HIVE-20944: Not validate stats during query compilation
Backport	HIVE-20940: Bridge cases in which Calcite's type resolution is more stricter than Hive.
Backport	HIVE-20937: Postgres jdbc query fail with "LIMIT must not be negative"
Backport	HIVE-20926: Semi join reduction hint fails when bloom filter entries are high or when there are no stats
Backport	HIVE-20920: Use SQL constraints to improve join reordering algorithm

Type	Description
Backport	HIVE-20918: Flag to enable/disable pushdown of computation from Calcite into JDBC connection
Backport	HIVE-20915: Make dynamic sort partition optimization available to HoS and MR
Backport	HIVE-20910: Insert in bucketed table fails due to dynamic partition sort optimization
Backport	HIVE-20899: Keytab URI for LLAP YARN Service is restrictive to support HDFS only
Backport	HIVE-20898: For time related functions arguments may not be casted to a non nullable type
Backport	HIVE-20881: Constant propagation oversimplifies projections
Backport	HIVE-20880: Update default value for hive.stats.filter.in.min.ratio
Backport	HIVE-20873: Use Murmur hash for VectorHashKeyWrapperTwoLong to reduce hash collision
Backport	HIVE-20868: SMB Join fails intermittently when TezDummyOperator has child op in getFinalOp in MapRecordProcessor
Backport	HIVE-20853: Expose ShuffleHandler.registerDag in the llap daemon API
Backport	HIVE-20850: Push case conditional from projections to dimension tables if possible

Type	Description
Backport	HIVE-20842: Fix logic introduced in HIVE-20660 to estimate statistics for group by
Backport	HIVE-20839: "Cannot find field" error during dynamically partitioned hash join
Backport	HIVE-20835: Interaction between constraints and MV rewriting may create loop in Calcite planner
Backport	HIVE-20834: Hive QueryResultCache entries keeping reference to SemanticAnalyzer from cached query
Backport	HIVE-20830: JdbcStorageHandler range query assertion failure in some cases
Backport	HIVE-20829: JdbcStorageHandler range split throws NPE
Backport	HIVE-20827: Inconsistent results for empty arrays
Backport	HIVE-20826: Enhance HiveSemiJoin rule to convert join + group by on left side to Left Semi Join
Backport	HIVE-20821: Rewrite SUM0 into SUM + COALESCE combination
Backport	HIVE-20815: JdbcRecordReader.next shall not eat exception
Backport	HIVE-20813: udf to_epoch_milli need to support timestamp without time zone as well.

Type	Description
Backport	HIVE-20804: Further improvements to group by optimization with constraints
Backport	HIVE-20792: Inserting timestamp with zones truncates the data
Backport	HIVE-20788: Extended SJ reduction may backtrack columns incorrectly when creating filters
Backport	HIVE-20778: Join reordering may not be triggered if all joins in plan are created by decorrelation logic
Backport	HIVE-20772: record per-task CPU counters in LLAP
Backport	HIVE-20768: Adding Tumbling Window UDF
Backport	HIVE-20767: Multiple project between join operators may affect join reordering using constraints
Backport	HIVE-20762: NOTIFICATION_LOG cleanup interval is hardcoded as 60s and is too small
Backport	HIVE-20761: Select for update on notificat ion_sequence table has retry interval and retries count too small
Backport	HIVE-20751: Upgrade arrow version to 0.10.0
Backport	HIVE-20746: HiveProtoHookLogger does not close file at end of day.
Backport	HIVE-20744: Use SQL constraints to improve join reordering algorithm

Type	Description
Backport	HIVE-20740: Remove global lock in ObjectStore.setConf method. This cherrypick backports HIVE-20740 intended for Hive 3.2 and 4.x to 3.1.x
Backport	HIVE-20734: Beeline: When beeline-site.xml is and hive CLI redirects to beeline, it should use the system username/dummy password instead of prompting for one
Backport	HIVE-20731: keystore file in JdbcStorageHandler should be authorized
Backport	HIVE-20720: Add partition column option to JDBC handler
Backport	HIVE-20719: SELECT statement fails after UPDATE with hive.optimize.sort.dynamic.partition optimization and vectorization on
Backport	HIVE-20718: Add perf cli driver with constraints
Backport	HIVE-20716: Set default value for hive.cbo.stats.correlated.multi.key.joins to true
Backport	HIVE-20712: HivePointLookupOptimizer should extract deep cases
Backport	HIVE-20710: Constant folding may not create null constants without types
Backport	HIVE-20706: external_jdbc_table2.q failing intermittently
Backport	HIVE-20704: Extend HivePreFilteringRule to support other functions

Type	Description
Backport	HIVE-20703: Put dynamic sort partition optimization under cost based decision
Backport	HIVE-20702: Account for overhead from datastructure aware estimations during mapjoin selection
Backport	HIVE-20692: Enable folding of NOT x IS (NOT) [TRUE|FALSE] expressions
Backport	HIVE-20691: Fix org.apache.hadoop.hive.cli. TestMiniLlapCliDriver.testCliDriver[cttl]
Backport	HIVE-20682: Async query execution can potentially fail if shared sessionHive is closed by master thread
Backport	HIVE-20676: HiveServer2: PrivilegeSynchroni zer is not set to daemon status
Backport	HIVE-20660: Group by statistics estimatio n could be improved by bounding the total number of rows to source table
Backport	HIVE-20652: JdbcStorageHandler push join of two different datasource to jdbc driver
Backport	HIVE-20651: JdbcStorageHandler password should be encrypted
Backport	HIVE-20649: LLAP aware memory manager for Orc writers
Backport	HIVE-20648: LLAP: Vector group by operator should use memory per executor

Type	Description
Backport	HIVE-20646: Partition filter condition is not pushed down to metastore query if it has IS NOT NULL
Backport	HIVE-20644: Avoid exposing sensitive infomation through a Hive Runtime exception
Backport	HIVE-20636: Improve number of null values estimation after outer join
Backport	HIVE-20632: Query with get_splits UDF fails if materialized view is created on queried table
Backport	HIVE-20627: Concurrent async queries intermittently fails with LockException and cause memory leak
Backport	HIVE-20623: Shared work: Extend sharing of map-join cache entries in LLAP
Backport	HIVE-20619: Include MultiDelimitSerDe in HiveServer2 By Default
Backport	HIVE-20618: During join selection BucketMap Join might be choosen for non bucketed tables
Backport	HIVE-20617: Fix type of constants in IN expressions to have correct type
Backport	HIVE-20612: Create new join multi-key correlation flag for CBO
Backport	HIVE-20603: "Wrong FS" error when inserting to partition after changing table location filesystem

Type	Description
Backport	HIVE-20601: EnvironmentContext null in ALTER_PARTITION event in DbNotificationList ener
Backport	HIVE-20583: Use canonical hostname only for kerberos auth in HiveConnection
Backport	HIVE-20582: Make hflush in hive proto logging configurable
Backport	HIVE-20563: Vectorization: CASE WHEN expression fails when THEN/ELSE type and result type are different
Backport	HIVE-20558: Change default of hive.hash table.key.count.adjustment to 0.99
Backport	HIVE-20552: Get Schema from LogicalPlan faster
Backport	HIVE-20550: Switch WebHCat to use beeline to submit Hive queries
Backport	HIVE-20537: Multi-column joins estimates with uncorrelated columns different in CBO and Hive
Backport	HIVE-20524: Schema Evolution checking is broken in going from Hive version 2 to version 3 for ALTER TABLE VARCHAR to DECIMAL
Backport	HIVE-20522: HiveFilterSetOpTransposeRule may throw assertion error due to nullability of fields

Type	Description
Backport	HIVE-20521: HS2 doAs=true has permission issue with hadoop.tmp.dir, with MR and S3A filesystem
Backport	HIVE-20515: Empty query results when using results cache and query temp dir, results cache dir in different filesystems
Backport	HIVE-20508: Hive does not support user names of type "user@realm"
Backport	HIVE-20507: Beeline: Add a utility command to retrieve all uris from beeline-site.xml
Backport	HIVE-20505: upgrade org.openjdk.jmh:jmh-core to 1.21
Backport	HIVE-20503: Use datastructure aware estimations during mapjoin selection
Backport	HIVE-20498: Support date type for column stats autogather
Backport	HIVE-20496: Vectorization: Vectorized PTF IllegalStateException
Backport	HIVE-20494: GenericUDFRestrictInformationSchema is broken after HIVE-19440
Backport	HIVE-20477: OptimizedSql is not shown if the expression contains INs
Backport	HIVE-20467: Allow IF NOT EXISTS/IF EXISTS in Resource plan creation/drop
Backport	HIVE-20462: "CREATE VIEW IF NOT EXISTS" fails if view already exists

Type	Description
Backport	HIVE-20455: Log spew from security.authorization.PrivilegeSynchonizer.run
Backport	HIVE-20439: Use the inflated memory limit during join selection for llap
Backport	HIVE-20433: Implicit String to Timestamp conversion is slow
Backport	HIVE-20432: Rewrite BETWEEN to IN for integer types for stats estimation
Backport	HIVE-20423: Set NULLS LAST as the default null ordering
Backport	HIVE-20418: LLAP IO may not handle ORC files that have row index disabled correctly for queries with no columns selected
Backport	HIVE-20412: NPE in HiveMetaHook
Backport	HIVE-20406: Nested Coalesce giving incorrect results
Backport	HIVE-20399: CTAS w/a custom table location that is not fully qualified fails for MM tables
Backport	HIVE-20393: Semijoin Reduction : markSemiJoinForDPP behaves inconsistently
Backport	HIVE-20391: HiveAggregateReduceFunctionsRule may infer wrong return type when decomposing aggregate function
Backport	HIVE-20383: Invalid queue name and synchronisation issues in hive proto events hook.

Type	Description
Backport	HIVE-20367: Vectorization: Support streaming for PTF AVG, MAX, MIN, SUM
Backport	HIVE-20366: TPC-DS query78 stats estimates are off for is null filte
Backport	HIVE-20364: Update default for hive.map.aggr.hash.min.reduction
Backport	HIVE-20352: Vectorization: Support grouping function
Backport	HIVE-20347: hive.optimize.sort.dynamic.partition should work with partitioned CTAS and MV
Backport	HIVE-20345: Drop database may hang if the tables get deleted from a different call
Backport	HIVE-20343: Hive 3: CTAS does not respect transactional_properties
Backport	HIVE-20340: Druid Needs Explicit CASTs from Timestamp to STRING when the output of timestamp function is used as Strin
Backport	HIVE-20339: Vectorization: Lift unneeded restriction causing some PTF with RANK not to be vectorized
Backport	HIVE-20337: CachedStore: getPartitionsByExpr is not populating the partition list correctly
Backport	HIVE-20336: Masking and filtering policies for materialized views

Type	Description
Backport	HIVE-20326: Create constraints with RELY as default instead of NO RELY
Backport	HIVE-20321: Vectorization: Cut down memory size of 1 col VectorHashKeyWrapper to <1 CacheLine
Backport	HIVE-20320: Turn on hive.optimize.remove.sq_count_check flag
Backport	HIVE-20315: Vectorization: Fix more NULL / Wrong Results issues and avoid unnecessary casts/conversions
Backport	HIVE-20314: Include partition pruning in materialized view rewriting
Backport	HIVE-20312: Allow arrow clients to use their own BufferAllocator with LlapOutputFormatService
Backport	HIVE-20302: LLAP: non-vectorized execution in IO ignores virtual columns, including ROW__ID
Backport	HIVE-20300: VectorFileSinkArrowOperator
Backport	HIVE-20299: potential race in LLAP signer unit test
Backport	HIVE-20296: Improve HivePointLookupOptimizerRule to be able to extract from more sophisticated contexts
Backport	HIVE-20294: Vectorization: Fix NULL / Wrong Results issues in COALESCE / ELT

Type	Description
Backport	HIVE-20292: Bad join ordering in tpcds query93 with primary constraint defined
Backport	HIVE-20290: Lazy initialize ArrowColumnarBatchSerDe so it doesn't allocate buffers during GetSplits
Backport	HIVE-20281: SharedWorkOptimizer fails with 'operator cache contents and actual plan differ'
Backport	HIVE-20277: Vectorization: Case expressions that return BOOLEAN are not supported for FILTER
Backport	HIVE-20267: Expanding WebUI to include form to dynamically config log levels
Backport	HIVE-20263: Typo in HiveReduceExpressionsWithStatsRule variable
Backport	HIVE-20260: NDV of a column shouldn't be scaled when row count is changed by filter on another column
Backport	HIVE-20252: Semijoin Reduction : Cycles due to semi join branch may remain undetected if small table side has a map join upstream.
Backport	HIVE-20245: Vectorization: Fix NULL / Wrong Results issues in BETWEEN / IN
Backport	HIVE-20241: Support partitioning spec in CTAS statements
Backport	HIVE-20240: Semijoin Reduction: Use local variable to check for external table condition

Type	Description
Backport	HIVE-20226: HMS getNextNotification will throw exception when request maxEvents exceed table's max_rows
Backport	HIVE-20225: SerDe to support Teradata Binary Format
Backport	HIVE-20213: Upgrade Calcite to 1.17.0
Backport	HIVE-20212: Hiveserver2 in http mode emitting metric default.General.open_connec tions incorrectly
Backport	HIVE-20210: Simple Fetch optimizer should lead to MapReduce when filter on non-parti tion column and conversion is minimal
Backport	HIVE-20209: Metastore connection fails for first attempt in repl dump
Backport	HIVE-20207: Vectorization: Fix NULL / Wrong Results issues in Filter / Compare
Backport	HIVE-20204: Type conversion during IN
Backport	HIVE-20203: Arrow SerDe leaks a DirectByt eBuffer
Backport	HIVE-20197: Vectorization: Add DECIMAL_6 4 testing, add Date/Interval/Timestamp arithmetic, and add more GROUP BY Aggregati on
Backport	HIVE-20193: cboInfo is not present in the explain plan json

Type	Description
Backport	HIVE-20192: HS2 with embedded metastore is leaking JDOPersistenceManager objects
Backport	HIVE-20183: Inserting from bucketed table can cause data loss, if the source table contains empty bucket
Backport	HIVE-20177: Vectorization: Reduce KeyWrapper allocation in GroupBy Streaming mode
Backport	HIVE-20174: Vectorization: Fix NULL / Wrong Results issues in GROUP BY Aggregation Functions
Backport	HIVE-20172: StatsUpdater failed with GSS Exception while trying to connect to remote metastore
Backport	HIVE-20153: Count and Sum UDF consume more memory in Hive 2+
Backport	HIVE-20152: reset db state, when repl dump fails, so rename table can be done
Backport	HIVE-20149: TestHiveCli failing/timing out
Backport	HIVE-20130: Better logging for information schema synchronizer
Backport	HIVE-20129: Revert to position based schema evolution for orc tables
Backport	HIVE-20118: SessionStateUserAuthenticat or.getGroupNames
Backport	HIVE-20116: TezTask is using parent logger

Type	Description
Backport	HIVE-20115: Acid tables should not use footer scan for analyze
Backport	HIVE-20103: WM: Only Aggregate DAG counters if at least one is used
Backport	HIVE-20101: BloomKFilter: Avoid using the local byte[] arrays entirely
Backport	HIVE-20100: OpTraits : Select Optraits should stop when a mismatch is detected
Backport	HIVE-20098: Statistics: NPE when getting Date column partition statistics
Backport	HIVE-20095: Fix feature to push computation to jdbc external tables
Backport	HIVE-20093: LlapOutputFomatService: Use ArrowBuf with Netty for Accounting
Backport	HIVE-20090: Extend creation of semijoin reduction filters to be able to discover new opportunities
Backport	HIVE-20088: Beeline config location path is assembled incorrectly
Backport	HIVE-20082: HiveDecimal to string conversion doesn't format the decimal correctly
Backport	HIVE-20069: Fix reoptimization in case of DPP and Semijoin optimization
Backport	HIVE-20051: Skip authorization for temp tables

Type	Description
Backport	HIVE-20044: Arrow Serde should pad char values and handle empty strings correctly
Backport	HIVE-20028: Metastore client cache config is used incorrectly
Backport	HIVE-20025: Clean-up of event files created by HiveProtoLoggingHook
Backport	HIVE-20020: Hive contrib jar should not be in lib
Backport	HIVE-20013: Add an Implicit cast to date type for to_date function
Backport	HIVE-20011: Move away from append mode in proto logging hook
Backport	HIVE-20005: acid_table_stats, acid_no_b uckets, etc - query result change on the branch
Backport	HIVE-20004: Wrong scale used by ConvertDe cimal64ToDecimal results in incorrect results
Backport	HIVE-19995: Aggregate row traffic for acid tables
Backport	HIVE-19993: Using a table alias which also appears as a column name is not possible
Backport	HIVE-19992: Vectorization: Follow-on to HIVE-19951 --> add call to SchemaEvo lution.isOnlyImplicitConversion to disable encoded LLAP I/O for ORC only when data type conversion is not implicit

Type	Description
Backport	HIVE-19989: Metastore uses wrong applicati on name for HADOOP2 metrics
Backport	HIVE-19981: Managed tables converted to external tables by the HiveStrictManagedM igration utility should be set to delete data when the table is dropped
Backport	HIVE-19967: SMB Join : Need Optraits for PTFOperator ala GBY Op
Backport	HIVE-19935: Hive WM session killed: Failed to update LLAP tasks count
Backport	HIVE-19924: Tag distcp jobs run by Repl Load
Backport	HIVE-19891: inserting into external tables with custom partition directories may cause data loss
Backport	HIVE-19850: Dynamic partition pruning in Tez is leading to 'No work found for tablescan' error
Backport	HIVE-19806: Sort qtests output to avoid flakiness in test results
Backport	HIVE-19770: Support for CBO for queries with multiple same columns in select
Backport	HIVE-19769: Create dedicated objects for DB and Table names
Backport	HIVE-19765: Add Parquet specific tests to BlobstoreCliDriver

Type	Description
Backport	HIVE-19759: Flaky test: TestRpc#testServer Port
Backport	HIVE-19711: Refactor Hive Schema Tool
Backport	HIVE-19701: getDelegationTokenFromMetaS tore doesn't need to be synchronized
Backport	HIVE-19694: Create Materialized View statement should check for MV name conflicts before running MV's SQL statement.
Backport	HIVE-19674: Group by Decimal Constants push down to Druid table
Backport	HIVE-19668: Over 30% of the heap wasted by duplicate org.antlr.runtime.CommonToken's and duplicate strings
Backport	HIVE-19663: refactor LLAP IO report generation
Backport	HIVE-19661: switch Hive UDFs to use Re2J regex engine
Backport	HIVE-19628: possible NPE in LLAP testSigning
Backport	HIVE-19568: Active/Passive HS2 HA: Disallow direct connection to passive HS2 instance
Backport	HIVE-19564: Vectorization: Fix NULL / Wrong Results issues in Arithmetic
Backport	HIVE-19552: Enable TestMiniDruidKafka CliDriver#druidkafkamini_basic.q
Backport	HIVE-19432: GetTablesOperation is too slow if the hive has too many databases and tables

Type	Description
Backport	HIVE-19360: CBO: Add an "optimizedSQL" to QueryPlan object
Backport	HIVE-19326: stats auto gather: incorrect aggregation during UNION queries
Backport	HIVE-19313: TestJdbcWithDBTokenStoreNoDoAs tests are failing
Backport	HIVE-19285: Add logs to the subclasses of MetaDataOperation
Backport	HIVE-19235: Update golden files for Minimr tests
Backport	HIVE-19104: When test MetaStore is started with retry the instances should be independent
Backport	HIVE-18986: Table rename will run java.lang.StackOverflowError in dataNucleus if the table contains large number of columns
Backport	HIVE-18920: CBO: Initialize the Janino providers ahead of 1st query
Backport	HIVE-18873: Skipping predicate pushdown for MR silently at HiveInputFormat can cause storage handlers to produce erroneous result
Backport	HIVE-18871: hive on tez execution error due to set hive.aux.jars.path to hdfs://
Backport	HIVE-18725: Improve error handling for subqueries if there is wrong column reference

Type	Description
Backport	HIVE-18696: The partition folders might not get cleaned up properly in the HiveMetaS tore.add_partitions_core method if an
Backport	HIVE-18453: ACID: Add "CREATE TRANSACTI ONAL TABLE" syntax to unify ACID ORC & Parquet support
Backport	HIVE-18201: Disable XPROD_EDGE for sq_count_chec
Backport	HIVE-18140: Partitioned tables statistics can go wrong in basic stats mixed case
Backport	HIVE-17921: Aggregation with struct in LLAP produces wrong result
Backport	HIVE-17896: TopNKey: Create a standalone vectorizable TopNKey operator
Backport	HIVE-17840: HiveMetaStore eats exception if transactionalListeners.notifyEvent fail
Backport	HIVE-17043: Remove non unique columns from group by keys if not referenced later
Backport	HIVE-17040: Join elimination in the presence of FK relationship
Backport	HIVE-16839: Unbalanced calls to openTrans action/commitTransaction when alter the same partition concurrently
Backport	HIVE-16100: Dynamic Sorted Partition optimizer loses sibling operators

Type	Description
Backport	HIVE-15956: StackOverflowError when drop lots of partitions
Backport	HIVE-15177: Authentication with hive fails when kerberos auth type is set to fromSubject and principal contains _HOST
Backport	HIVE-14898: HS2 shouldn't log callstack for an empty auth header error
Backport	HIVE-14493: Partitioning support for materiali zed views
Backport	HIVE-14431: Recognize COALESCE as CASE
Backport	HIVE-13457: Create HS2 REST API endpoints for monitoring information
Backport	HIVE-12342: Set default value of hive.opti mize.index.filter to true
Backport	HIVE-10296: Cast exception observed when hive runs a multi join query on metastore
Backport	HIVE-6980: Drop table by using direct sql

Amazon EMR 6.6.0 - Hive configuration changes

- As part of OSS change HIVE-20703, the property to sort dynamic partitions, `hive.optimize.sort.dynamic.partition`, has been replaced with `hive.optimize.sort.dynamic.partition.threshold`.

  The `hive.optimize.sort.dynamic.partition.threshold` configuration has the following potential values:

Value	Description
`0` *(default)*	Makes the optimization to sort dynamic partitions a cost-based decision when ORC files are used. The max number of writers allowed in INSERT queries is computed based on (executor/container memory) * (percentage of memory taken by orc) divided by max memory (stripe size) taken by a single writer.
`-1`	Disables the optimization to sort dynamic partitions completely.
1	Enables global sorting of dynamic partitions. This keeps only one record writer open for each partition value in the reducer, thereby reducing the memory pressure on reducers.
2 *(or greater integer)*	Tells Hive to use the specified integer as threshold for the maximum number of writers.

Amazon EMR 6.6.0 - Hive known issues

- Queries with windowing functions on the same column as join may lead to invalid transformations as reported in HIVE-25278 and cause incorrect results or query failures. As a workaround, you can disable CBO at the query level for such queries. Contact AWS support for further information.
- Amazon EMR 6.6.0 includes Hive software version 3.1.2. Hive 3.1.2 introduces a feature that splits text files if they contain a header and footer (HIVE-21924). The Apache Tez App Master reads each of your files to determine offset points in the data range. These behaviors combined could negatively impact performance if your queries read a large number of small text files. As a workaround, use `CombineHiveInputFormat` and tune the max split size by configuring the following properties:

```
SET hive.tez.input.format=org.apache.hadoop.hive.ql.io.CombineHiveInputFormat;
SET mapreduce.input.fileinputformat.split.maxsize=16777216;
```

- With Amazon EMR 6.6.0 through 6.9.x, INSERT queries with dynamic partition and an ORDER BY or SORT BY clause will always have two reducers. This issue is caused by OSS change

HIVE-20703, which puts dynamic sort partition optimization under cost-based decision. If your workload doesn't require sorting of dynamic partitions, we recommend that you set the `hive.optimize.sort.dynamic.partition.threshold` property to `-1` to disable the new feature and get the correctly calculated number of reducers. This issue is fixed in OSS Hive as part of HIVE-22269 and is fixed in Amazon EMR 6.10.0.

Hudi

Apache Hudi is an open-source data management framework used to simplify incremental data processing and data pipeline development by providing record-level insert, update, upsert, and delete capabilities. *Upsert* refers to the ability to insert records into an existing dataset if they do not already exist or to update them if they do. By efficiently managing how data is laid out in Amazon S3, Hudi allows data to be ingested and updated in near real time. Hudi carefully maintains metadata of the actions performed on the dataset to help ensure that the actions are atomic and consistent.

Hudi is integrated with Apache Spark, Apache Hive, and Presto. In Amazon EMR release versions 6.1.0 and later, Hudi is also integrated with Trino (PrestoSQL).

With Amazon EMR release version 5.28.0 and later, EMR installs Hudi components by default when Spark, Hive, Presto, or Flink are installed. You can use Spark or the Hudi DeltaStreamer utility to create or update Hudi datasets. You can use Hive, Spark, Presto, or Flink to query a Hudi dataset interactively or build data processing pipelines using *incremental pull*. Incremental pull refers to the ability to pull only the data that changed between two actions.

These features make Hudi suitable for the following use cases:

- Working with streaming data from sensors and other Internet of Things (IoT) devices that require specific data insertion and update events.
- Complying with data privacy regulations in applications where users might choose to be forgotten or modify their consent for how their data can be used.
- Implementing a change data capture (CDC) system that allows you to apply changes to a dataset over time.

The following table lists the version of Hudi included in the latest release of the Amazon EMR 7.x series, along with the components that Amazon EMR installs with Hudi.

For the version of components installed with Hudi in this release, see Release 7.2.0 Component Versions.

Hudi version information for emr-7.2.0

Amazon EMR Release Label	Hudi Version	Components Installed With Hudi
emr-7.2.0	Hudi 0.14.1-amzn-1	Not available.

The following table lists the version of Hudi included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with Hudi.

For the version of components installed with Hudi in this release, see Release 6.15.0 Component Versions.

Hudi version information for emr-6.15.0

Amazon EMR Release Label	Hudi Version	Components Installed With Hudi
emr-6.15.0	Hudi 0.14.0-amzn-0	Not available.

> **Note**
>
> Amazon EMR release 6.8.0 comes with Apache Hudi 0.11.1; however, Amazon EMR 6.8.0 clusters are also compatible with the open-source `hudi-spark3.3-bundle_2.12` from Hudi 0.12.0.

The following table lists the version of Hudi included in the latest release of the Amazon EMR 5.x series, along with the components that Amazon EMR installs with Hudi.

For the version of components installed with Hudi in this release, see Release 5.36.2 Component Versions.

Hudi version information for emr-5.36.2

Amazon EMR Release Label	Hudi Version	Components Installed With Hudi
emr-5.36.2	Hudi 0.10.1-amzn-1	Not available.

Topics

How Hudi works

When using Hudi with Amazon EMR, you can write data to the dataset using the Spark Data Source API or the Hudi DeltaStreamer utility. Hudi organizes a dataset into a partitioned directory structure under a *`basepath`* that is similar to a traditional Hive table. The specifics of how the data is laid out as files in these directories depend on the dataset type that you choose. You can choose either Copy on Write (CoW) or Merge on Read (MoR).

Regardless of the dataset type, each partition in a dataset is uniquely identified by its `partitionpath` relative to the `basepath`. Within each partition, records are distributed into multiple data files. For more information, see File management in the Apache Hudi documentation.

Each action in Hudi has a corresponding commit, identified by a monotonically increasing timestamp known as an *Instant*. Hudi keeps a series of all actions performed on the dataset as a timeline. Hudi relies on the timeline to provide snapshot isolation between readers and writers, and to enable roll back to a previous point in time. For more information about the actions that Hudi records and the state of actions, see Timeline in the Apache Hudi documentation.

Understanding dataset storage types: Copy on write vs. merge on read

When you create a Hudi dataset, you specify that the dataset is either copy on write or merge on read.

- **Copy on Write (CoW) –** Data is stored in a columnar format (Parquet), and each update creates a new version of files during a write. CoW is the default storage type.
- **Merge on Read (MoR) –** Data is stored using a combination of columnar (Parquet) and row-based (Avro) formats. Updates are logged to row-based *delta* files and are compacted as needed to create new versions of the columnar files.

With CoW datasets, each time there is an update to a record, the file that contains the record is rewritten with the updated values. With a MoR dataset, each time there is an update, Hudi writes only the row for the changed record. MoR is better suited for write- or change-heavy workloads with fewer reads. CoW is better suited for read-heavy workloads on data that changes less frequently.

Hudi provides three logical views for accessing the data:

- **Read-optimized view –** Provides the latest committed dataset from CoW tables and the latest compacted dataset from MoR tables.
- **Incremental view –** Provides a change stream between two actions out of a CoW dataset to feed downstream jobs and extract, transform, load (ETL) workflows.
- **Real-time view –** Provides the latest committed data from a MoR table by merging the columnar and row-based files inline.

When you query the read-optimized view, the query returns all compacted data but does not include the latest delta commits. Querying this data provides good read performance but omits the freshest data. When you query the real-time view, Hudi merges the compacted data with the delta commits on read. The freshest data is available to query, but the computational overhead of merging makes the query less performant. The ability to query either compacted data or real-time data allows you to choose between performance and flexibility when you query.

For more information about the tradeoffs between storage types, see Storage types & views in Apache Hudi documentation.

Hudi creates two tables in the Hive metastore for MoR: a table with the name that you specified, which is a read-optimized view, and a table with the same name appended with `_rt`, which is a real-time view. You can query both tables.

Registering a Hudi dataset with your metastore

When you register a Hudi table with the Hive metastore, you can query Hudi tables using Hive, Spark SQL or Presto as you would any other table. In addition, you can integrate Hudi with AWS Glue by configuring Hive and Spark to use the AWS Glue Data Catalog as the metastore. For MoR tables, Hudi registers the dataset as two tables in the Metastore: a table with the name that you specified, which is a read-optimized view, and a table with the same name appended with `_rt`, which is a real-time view.

You register a Hudi table with the Hive metastore when you use Spark to create a Hudi dataset by setting the `HIVE_SYNC_ENABLED_OPT_KEY` option to `"true"` and providing other required properties. For more information, see Work with a Hudi dataset. In addition, you can use the hive_sync_tool command line utility to register a Hudi dataset as a table in your metastore, separately.

Considerations and limitations for using Hudi on Amazon EMR

- **Record key field cannot be null or empty** – The field that you specify as the record key field cannot have `null` or empty values.
- **Schema updated by default on upsert and insert** – Hudi provides an interface, `HoodieRecordPayload` that determines how the input DataFrame and existing Hudi dataset are merged to produce a new, updated dataset. Hudi provides a default implementation of this class, `OverwriteWithLatestAvroPayload`, that overwrites existing records and updates the schema as specified in the input DataFrame. To customize this logic for implementing merge and partial updates, you can provide an implementation of the `HoodieRecordPayload` interface using the `DataSourceWriteOptions.PAYLOAD_CLASS_OPT_KEY` parameter.
- **Deletion requires schema** – When deleting, you must specify the record key, the partition key, and the pre-combine key fields. Other columns can be made `null` or empty, but the full schema is required.
- **MoR table limitations** – MoR tables do not support savepointing. You can query MoR tables using the read-optimized view or the real-time view (*`tableName`*`_rt`) from Spark SQL, Presto, or Hive. Using the read-optimized view only exposes base file data, and does not expose a merged view of base and log data.
- **Hive**
  - For registering tables in the Hive metastore, Hudi expects the Hive Thrift server to be running at the default port `10000`. If you override this port with a custom port, pass the `HIVE_URL_OPT_KEY` option as shown in the following example.

    ```
    .option(DataSourceWriteOptions.HIVE_URL_OPT_KEY, "jdbc:hive2://localhost:override-port-number
    ```

  - The `timestamp` data type in Spark is registerd as `long` data type in Hive, and not as Hive's `timestamp` type.
- **Presto**
  - Presto does not support reading MoR real time tables in Hudi versions below 0.6.0.

- Presto only supports snapshot queries.
- For Presto to correctly interpret Hudi dataset columns, set the `hive.parquet_use_column_names` value to `true`.
  - To set the value for a session, in the Presto shell, run the following command:

    ```
    set session hive.parquet_use_column_names=true
    ```

  - To set the value at the cluster level, use the `presto-connector-hive` configuration classification to set `hive.parquet.use_column_names` to `true`, as shown in the following example. For more information, see Configure applications.

    ```
    [
      {
        "Classification": "presto-connector-hive",
        "Properties": {
          "hive.parquet.use-column-names": "true"
        }
      }
    ]
    ```

- **HBase Index**
  - The HBase version used to *build* Hudi might be different from what is listed in the EMR Release Guide. To pull in the correct dependencies for your Spark session, run the following command.

  ```
  spark-shell \
  --jars /usr/lib/spark/external/lib/spark-avro.jar,/usr/lib/hudi/cli/lib/*.jar \
  --conf "spark.serializer=org.apache.spark.serializer.KryoSerializer" \
  --conf "spark.sql.hive.convertMetastoreParquet=false"
  ```

Create a cluster with Hudi installed

With Amazon EMR release version 5.28.0 and later, Amazon EMR installs Hudi components by default when Spark, Hive, or Presto is installed. To use Hudi on Amazon EMR, create a cluster with one or more of the following applications installed:

- Hadoop
- Hive
- Spark

- Presto
- Flink

You can create a cluster using the AWS Management Console, the AWS CLI, or the Amazon EMR API.

To create a cluster with Hudi using the AWS Management Console

1. Navigate to the new Amazon EMR console and select **Switch to the old console** from the side navigation. For more information on what to expect when you switch to the old console, see Using the old console.
2. Choose **Create cluster**, **Go to advanced options**.
3. Under Software Configuration, choose **emr-5.28.0** or later for **Release** and select **Hadoop**, **Hive**, **Spark**, **Presto**, and **Tez** along with other applications that your cluster requires.
4. Configure other options as required for your application, and then choose **Next**.
5. Configure options for **Hardware** and **General cluster settings** as desired.
6. For **Security Options**, we recommend that you select an **EC2 key pair** that you can use to connect to the master node command line using SSH. This allows you to run the Spark shell commands, Hive CLI commands, and Hudi CLI commands described in this guide.
7. Choose other security options as desired, and then choose **Create cluster**.

Work with a Hudi dataset

Hudi supports inserting, updating, and deleting data in Hudi datasets through Spark. For more information, see Writing Hudi tables in Apache Hudi documentation.

The following examples demonstrate how to launch the interactive Spark shell, use Spark submit, or use Amazon EMR Notebooks to work with Hudi on Amazon EMR. You can also use the Hudi DeltaStreamer utility or other tools to write to a dataset. Throughout this section, the examples demonstrate working with datasets using the Spark shell while connected to the master node using SSH as the default `hadoop` user.

Launch the Spark shell using Amazon EMR 6.7 and later

When running `spark-shell`, `spark-submit`, or `spark-sql` using Amazon EMR 6.7.0 or later, pass the following commands.

> **Note**
>
> Amazon EMR 6.7.0 uses Apache Hudi 0.11.0-amzn-0, which contains significant improvements over previous Hudi versions. For more information, see the Apache Hudi 0.11.0 Migration Guide. The examples on this tab reflect these changes.

To open the Spark shell on the primary node

1. Connect to the primary node using SSH. For more information, see Connect to the primary node using SSH in the *Amazon EMR Management Guide*.
2. Enter the following command to launch the Spark shell. To use the PySpark shell, replace *spark-shell* with *pyspark*.

```
spark-shell --jars /usr/lib/hudi/hudi-spark-bundle.jar \
--conf "spark.serializer=org.apache.spark.serializer.KryoSerializer" \
--conf
 "spark.sql.catalog.spark_catalog=org.apache.spark.sql.hudi.catalog.HoodieCatalog"
 \
--conf "spark.sql.extensions=org.apache.spark.sql.hudi.HoodieSparkSessionExtension"
```

Launch the Spark shell using Amazon EMR 6.6 and earlier

When running `spark-shell`, `spark-submit`, or `spark-sql` using Amazon EMR 6.6.x or earlier, pass the following commands.

> **Note**
>
> - Amazon EMR 6.2 and 5.31 and later (Hudi 0.6.x and later) can omit the `spark-avro.jar` from the configuration.
> - Amazon EMR 6.5 and 5.35 and later (Hudi 0.9.x and later) can omit `spark.sql.hive.convertMetastoreParquet=false` from the configuration.
> - Amazon EMR 6.6 and 5.36 and later (Hudi 0.10.x and later) must include the `HoodieSparkSessionExtension` config as described in the Version: 0.10.0 Spark Guide:

```
--conf
 "spark.sql.extensions=org.apache.spark.sql.hudi.HoodieSparkSessionExtension"
 \
```

To open the Spark shell on the primary node

1. Connect to the primary node using SSH. For more information, see Connect to the primary node using SSH in the *Amazon EMR Management Guide*.
2. Enter the following command to launch the Spark shell. To use the PySpark shell, replace *spark-shell* with *pyspark*.

```
spark-shell \
--conf "spark.serializer=org.apache.spark.serializer.KryoSerializer" \
--conf "spark.sql.hive.convertMetastoreParquet=false" \
--jars /usr/lib/hudi/hudi-spark-bundle.jar,/usr/lib/spark/external/lib/spark-
avro.jar
```

Use Hudi with Amazon EMR Notebooks using Amazon EMR 6.7 and later

To use Hudi with Amazon EMR Notebooks, you must first copy the Hudi jar files from the local file system to HDFS on the master node of the notebook cluster. You then use the notebook editor to configure your EMR notebook to use Hudi.

To use Hudi with Amazon EMR Notebooks

1. Create and launch a cluster for Amazon EMR Notebooks. For more information, see Creating Amazon EMR clusters for notebooks in the *Amazon EMR Management Guide*.
2. Connect to the master node of the cluster using SSH and then copy the jar files from the local filesystem to HDFS as shown in the following examples. In the example, we create a directory in HDFS for clarity of file management. You can choose your own destination in HDFS, if desired.

```
hdfs dfs -mkdir -p /apps/hudi/lib
```

```
hdfs dfs -copyFromLocal /usr/lib/hudi/hudi-spark-bundle.jar /apps/hudi/lib/hudi-
spark-bundle.jar
```

3. Open the notebook editor, enter the code from the following example, and run it.

```
%%configure
{ "conf": {
            "spark.jars":"hdfs:///apps/hudi/lib/hudi-spark-bundle.jar",
            "spark.serializer":"org.apache.spark.serializer.KryoSerializer",
            "spark.sql.catalog.spark_catalog":
 "org.apache.spark.sql.hudi.catalog.HoodieCatalog",

 "spark.sql.extensions":"org.apache.spark.sql.hudi.HoodieSparkSessionExtension"
          }}
```

Use Hudi with Amazon EMR Notebooks using Amazon EMR 6.6 and earlier

To use Hudi with Amazon EMR Notebooks, you must first copy the Hudi jar files from the local file system to HDFS on the master node of the notebook cluster. You then use the notebook editor to configure your EMR notebook to use Hudi.

To use Hudi with Amazon EMR Notebooks

1. Create and launch a cluster for Amazon EMR Notebooks. For more information, see Creating Amazon EMR clusters for notebooks in the *Amazon EMR Management Guide*.
2. Connect to the master node of the cluster using SSH and then copy the jar files from the local filesystem to HDFS as shown in the following examples. In the example, we create a directory in HDFS for clarity of file management. You can choose your own destination in HDFS, if desired.

```
hdfs dfs -mkdir -p /apps/hudi/lib
```

```
hdfs dfs -copyFromLocal /usr/lib/hudi/hudi-spark-bundle.jar /apps/hudi/lib/hudi-
spark-bundle.jar
```

```
hdfs dfs -copyFromLocal /usr/lib/spark/external/lib/spark-avro.jar /apps/hudi/lib/
spark-avro.jar
```

3. Open the notebook editor, enter the code from the following example, and run it.

```
{ "conf": {
            "spark.jars":"hdfs:///apps/hudi/lib/hudi-spark-bundle.jar,hdfs:///apps/
hudi/lib/spark-avro.jar",
            "spark.serializer":"org.apache.spark.serializer.KryoSerializer",
            "spark.sql.hive.convertMetastoreParquet":"false"
          }}
```

Initialize a Spark session for Hudi

When you use Scala, you must import the following classes in your Spark session. This needs to be done once per Spark session.

```
import org.apache.spark.sql.SaveMode
import org.apache.spark.sql.functions._
import org.apache.hudi.DataSourceWriteOptions
import org.apache.hudi.DataSourceReadOptions
import org.apache.hudi.config.HoodieWriteConfig
import org.apache.hudi.hive.MultiPartKeysValueExtractor
import org.apache.hudi.hive.HiveSyncConfig
import org.apache.hudi.sync.common.HoodieSyncConfig
```

Write to a Hudi dataset

The following examples show how to create a DataFrame and write it as a Hudi dataset.

> **Note**
>
> To paste code samples into the Spark shell, type **`:paste`** at the prompt, paste the example, and then press **`CTRL`** + **`D`**.

Each time you write a DataFrame to a Hudi dataset, you must specify `DataSourceWriteOptions`. Many of these options are likely to be identical between write operations. The following example specifies common options using the *`hudiOptions`* variable, which subsequent examples use.

Write using Scala with Amazon EMR 6.7 and later

Note

Amazon EMR 6.7.0 uses Apache Hudi 0.11.0-amzn-0, which contains significant improvements over previous Hudi versions. For more information, see the Apache Hudi 0.11.0 Migration Guide. The examples on this tab reflect these changes.

```
// Create a DataFrame
val inputDF = Seq(
 ("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
 ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
 ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
 ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z"),
 ("104", "2015-01-02", "2015-01-01T12:15:00.512679Z"),
 ("105", "2015-01-02", "2015-01-01T13:51:42.248818Z")
 ).toDF("id", "creation_date", "last_update_time")

//Specify common DataSourceWriteOptions in the single hudiOptions variable
val hudiOptions = Map[String,String](
  HoodieWriteConfig.TBL_NAME.key -> "tableName",
  DataSourceWriteOptions.TABLE_TYPE.key -> "COPY_ON_WRITE",
  DataSourceWriteOptions.RECORDKEY_FIELD_OPT_KEY -> "id",
  DataSourceWriteOptions.PARTITIONPATH_FIELD_OPT_KEY -> "creation_date",
  DataSourceWriteOptions.PRECOMBINE_FIELD_OPT_KEY -> "last_update_time",
  DataSourceWriteOptions.HIVE_SYNC_ENABLED_OPT_KEY -> "true",
  DataSourceWriteOptions.HIVE_TABLE_OPT_KEY -> "tableName",
  DataSourceWriteOptions.HIVE_PARTITION_FIELDS_OPT_KEY -> "creation_date",
  HoodieSyncConfig.META_SYNC_PARTITION_EXTRACTOR_CLASS.key ->
 "org.apache.hudi.hive.MultiPartKeysValueExtractor",
  HoodieSyncConfig.META_SYNC_ENABLED.key -> "true",
  HiveSyncConfig.HIVE_SYNC_MODE.key -> "hms",
  HoodieSyncConfig.META_SYNC_TABLE_NAME.key -> "tableName",
  HoodieSyncConfig.META_SYNC_PARTITION_FIELDS.key -> "creation_date"
)

// Write the DataFrame as a Hudi dataset
(inputDF.write
    .format("hudi")
    .options(hudiOptions)
    .option(DataSourceWriteOptions.OPERATION_OPT_KEY,"insert")
```

```
    .mode(SaveMode.Overwrite)
    .save("s3://DOC-EXAMPLE-BUCKET/myhudidataset/"))
```

Write using Scala with Amazon EMR 6.6 and earlier

```
// Create a DataFrame
val inputDF = Seq(
 ("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
 ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
 ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
 ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z"),
 ("104", "2015-01-02", "2015-01-01T12:15:00.512679Z"),
 ("105", "2015-01-02", "2015-01-01T13:51:42.248818Z")
 ).toDF("id", "creation_date", "last_update_time")

//Specify common DataSourceWriteOptions in the single hudiOptions variable
val hudiOptions = Map[String,String](
  HoodieWriteConfig.TABLE_NAME -> "tableName",
  DataSourceWriteOptions.TABLE_TYPE_OPT_KEY -> "COPY_ON_WRITE",
  DataSourceWriteOptions.RECORDKEY_FIELD_OPT_KEY -> "id",
  DataSourceWriteOptions.PARTITIONPATH_FIELD_OPT_KEY -> "creation_date",
  DataSourceWriteOptions.PRECOMBINE_FIELD_OPT_KEY -> "last_update_time",
  DataSourceWriteOptions.HIVE_SYNC_ENABLED_OPT_KEY -> "true",
  DataSourceWriteOptions.HIVE_TABLE_OPT_KEY -> "tableName",
  DataSourceWriteOptions.HIVE_PARTITION_FIELDS_OPT_KEY -> "creation_date",
  DataSourceWriteOptions.HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY ->
 classOf[MultiPartKeysValueExtractor].getName
)

// Write the DataFrame as a Hudi dataset
(inputDF.write
    .format("org.apache.hudi")
    .option(DataSourceWriteOptions.OPERATION_OPT_KEY,
 DataSourceWriteOptions.INSERT_OPERATION_OPT_VAL)
    .options(hudiOptions)
    .mode(SaveMode.Overwrite)
    .save("s3://DOC-EXAMPLE-BUCKET/myhudidataset/"))
```

Write using PySpark

```
# Create a DataFrame
inputDF = spark.createDataFrame(
    [
```

```
        ("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
        ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
        ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
        ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z"),
        ("104", "2015-01-02", "2015-01-01T12:15:00.512679Z"),
        ("105", "2015-01-02", "2015-01-01T13:51:42.248818Z"),
    ],
    ["id", "creation_date", "last_update_time"]
)

# Specify common DataSourceWriteOptions in the single hudiOptions variable
hudiOptions = {
'hoodie.table.name': 'tableName',
'hoodie.datasource.write.recordkey.field': 'id',
'hoodie.datasource.write.partitionpath.field': 'creation_date',
'hoodie.datasource.write.precombine.field': 'last_update_time',
'hoodie.datasource.hive_sync.enable': 'true',
'hoodie.datasource.hive_sync.table': 'tableName',
'hoodie.datasource.hive_sync.partition_fields': 'creation_date',
'hoodie.datasource.hive_sync.partition_extractor_class':
 'org.apache.hudi.hive.MultiPartKeysValueExtractor'
}

# Write a DataFrame as a Hudi dataset
inputDF.write \
.format('org.apache.hudi') \
.option('hoodie.datasource.write.operation', 'insert') \
.options(**hudiOptions) \
.mode('overwrite') \
.save('s3://DOC-EXAMPLE-BUCKET/myhudidataset/')
```

Note

You might see "hoodie" instead of Hudi in code examples and notifications. The Hudi codebase widely uses the old "hoodie" spelling.

DataSourceWriteOptions reference for Hudi

Option	Description
TABLE_NAME	The table name under which to register the dataset.
TABLE_TYPE_OPT_KEY	Optional. Specifies whether the dataset is created as `"COPY_ON_WRITE"` or `"MERGE_ON_READ"` . The default is `"COPY_ON_WRITE"` .
RECORDKEY_FIELD_OPT_KEY	The record key field whose value will be used as the `recordKey` component of `HoodieKey` . Actual value will be obtained by invoking `.toString()` on the field value. Nested fields can be specified using the dot notation, for example, `a.b.c`.
PARTITIONPATH_FIELD_OPT_KEY	The partition path field whose value will be used as the `partitionPath` component of `HoodieKey` . The actual value will be obtained by invoking `.toString()` on the field value.
PRECOMBINE_FIELD_OPT_KEY	The field used in pre-combining before actual write. When two records have the same key value, Hudi picks the one with the largest value for the precombine field as determined by `Object.compareTo(..)` .

The following options are required only to register the Hudi dataset table in your metastore. If you do not register your Hudi dataset as a table in the Hive metastore, these options are not required.

DataSourceWriteOptions reference for Hive

Option	Description
HIVE_DATABASE_OPT_KEY	The Hive database to sync to. The default is `"default"` .
HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY	The class used to extract partition field values into Hive partition columns.
HIVE_PARTITION_FIELDS_OPT_KEY	The field in the dataset to use for determining Hive partition columns.
HIVE_SYNC_ENABLED_OPT_KEY	When set to `"true"`, registers the dataset with the Apache Hive metastore. The default is `"false"`.
HIVE_TABLE_OPT_KEY	Required. The name of the table in Hive to sync to. For example, `"my_hudi_table_cow"` .
HIVE_USER_OPT_KEY	Optional. The Hive user name to use when syncing. For example, `"hadoop"`.
HIVE_PASS_OPT_KEY	Optional. The Hive password for the user specified by `HIVE_USER_OPT_KEY` .
HIVE_URL_OPT_KEY	The Hive metastore URL.

Upsert data

The following example demonstrates how to upsert data by writing a DataFrame. Unlike the previous insert example, the `OPERATION_OPT_KEY` value is set to `UPSERT_OPERATION_OPT_VAL`. In addition, `.mode(SaveMode.Append)` is specified to indicate that the record should be appended.

Upsert using Scala with Amazon EMR 6.7 and later

Note

Amazon EMR 6.7.0 uses Apache Hudi 0.11.0-amzn-0, which contains significant improvements over previous Hudi versions. For more information, see the Apache Hudi 0.11.0 Migration Guide. The examples on this tab reflect these changes.

```
// Create a new DataFrame from the first row of inputDF with a different creation_date
 value
val updateDF = inputDF.limit(1).withColumn("creation_date", lit("new_value"))

(updateDF.write
    .format("hudi")
    .options(hudiOptions)
    .option(DataSourceWriteOptions.OPERATION_OPT_KEY, "upsert")
    .mode(SaveMode.Append)
    .save("s3://DOC-EXAMPLE-BUCKET/myhudidataset/"))
```

Upsert using Scala with Amazon EMR 6.6 and earlier

```
// Create a new DataFrame from the first row of inputDF with a different creation_date
 value
val updateDF = inputDF.limit(1).withColumn("creation_date", lit("new_value"))

(updateDF.write
    .format("org.apache.hudi")
    .option(DataSourceWriteOptions.OPERATION_OPT_KEY,
 DataSourceWriteOptions.UPSERT_OPERATION_OPT_VAL)
    .options(hudiOptions)
    .mode(SaveMode.Append)
    .save("s3://DOC-EXAMPLE-BUCKET/myhudidataset/"))
```

Upsert using PySpark

```
from pyspark.sql.functions import lit

# Create a new DataFrame from the first row of inputDF with a different creation_date
 value
```

```
updateDF = inputDF.limit(1).withColumn('creation_date', lit('new_value'))

updateDF.write \
    .format('org.apache.hudi') \
    .option('hoodie.datasource.write.operation', 'upsert') \
    .options(**hudiOptions) \
    .mode('append') \
    .save('s3://DOC-EXAMPLE-BUCKET/myhudidataset/')
```

Delete a record

To hard delete a record, you can upsert an empty payload. In this case, the `PAYLOAD_CLASS_OPT_KEY` option specifies the `EmptyHoodieRecordPayload` class. The example uses the same DataFrame, `updateDF`, used in the upsert example to specify the same record.

Delete using Scala with Amazon EMR 6.7 and later

> **Note**
>
> Amazon EMR 6.7.0 uses Apache Hudi 0.11.0-amzn-0, which contains significant improvements over previous Hudi versions. For more information, see the Apache Hudi 0.11.0 Migration Guide. The examples on this tab reflect these changes.

```
(updateDF.write
    .format("hudi")
    .options(hudiOptions)
    .option(DataSourceWriteOptions.OPERATION_OPT_KEY, "delete")
    .mode(SaveMode.Append)
    .save("s3://DOC-EXAMPLE-BUCKET/myhudidataset/"))
```

Delete using Scala with Amazon EMR 6.6 and earlier

```
(updateDF.write
    .format("org.apache.hudi")
    .option(DataSourceWriteOptions.OPERATION_OPT_KEY,
 DataSourceWriteOptions.UPSERT_OPERATION_OPT_VAL)
    .option(DataSourceWriteOptions.PAYLOAD_CLASS_OPT_KEY,
 "org.apache.hudi.common.model.EmptyHoodieRecordPayload")
```

```
    .mode(SaveMode.Append)
    .save("s3://DOC-EXAMPLE-BUCKET/myhudidataset/"))
```

Delete using PySpark

```
updateDF.write \
    .format('org.apache.hudi') \
    .option('hoodie.datasource.write.operation', 'upsert') \
    .option('hoodie.datasource.write.payload.class',
 'org.apache.hudi.common.model.EmptyHoodieRecordPayload') \
    .options(**hudiOptions) \
    .mode('append') \
    .save('s3://DOC-EXAMPLE-BUCKET/myhudidataset/')
```

You can also hard delete data by setting `OPERATION_OPT_KEY` to `DELETE_OPERATION_OPT_VAL` to remove all records in the dataset you submit. For instructions on performing soft deletes, and for more information about deleting data stored in Hudi tables, see Deletes in the Apache Hudi documentation.

Read from a Hudi dataset

To retrieve data at the present point in time, Hudi performs snapshot queries by default. Following is an example of querying the dataset written to S3 in Write to a Hudi dataset. Replace *`s3://DOC-EXAMPLE-BUCKET/myhudidataset`* with your table path, and add wildcard asterisks for each partition level, *plus one additional asterisk*. In this example, there is one partition level, so we've added two wildcard symbols.

Read using Scala with Amazon EMR 6.7 and later

> **Note**
>
> Amazon EMR 6.7.0 uses Apache Hudi 0.11.0-amzn-0, which contains significant improvements over previous Hudi versions. For more information, see the Apache Hudi 0.11.0 Migration Guide. The examples on this tab reflect these changes.

```
val snapshotQueryDF = spark.read
    .format("hudi")
    .load("s3://DOC-EXAMPLE-BUCKET/myhudidataset")
```

```
    .show()
```

Read using Scala with Amazon EMR 6.6 and earlier

```
(val snapshotQueryDF = spark.read
    .format("org.apache.hudi")
    .load("s3://DOC-EXAMPLE-BUCKET/myhudidataset" + "/*/*"))

snapshotQueryDF.show()
```

Read using PySpark

```
snapshotQueryDF = spark.read \
    .format('org.apache.hudi') \
    .load('s3://DOC-EXAMPLE-BUCKET/myhudidataset' + '/*/*')

snapshotQueryDF.show()
```

Incremental queries

You can also perform incremental queries with Hudi to get a stream of records that have changed since a given commit timestamp. To do so, set the `QUERY_TYPE_OPT_KEY` field to `QUERY_TYPE_INCREMENTAL_OPT_VAL`. Then, add a value for `BEGIN_INSTANTTIME_OPT_KEY` to obtain all records written since the specified time. Incremental queries are typically ten times more efficient than their batch counterparts since they only process changed records.

When you perform incremental queries, use the root (base) table path without the wildcard asterisks used for Snapshot queries.

> **Note**
>
> Presto does not support incremental queries.

Incremental queries using Scala

```
(val incQueryDF = spark.read
    .format("org.apache.hudi")
    .option(DataSourceReadOptions.QUERY_TYPE_OPT_KEY,
 DataSourceReadOptions.QUERY_TYPE_INCREMENTAL_OPT_VAL)
```

```
    .option(DataSourceReadOptions.BEGIN_INSTANTTIME_OPT_KEY, <beginInstantTime>)
    .load("s3://DOC-EXAMPLE-BUCKET/myhudidataset" ))

incQueryDF.show()
```

Incremental queries using PySpark

```
readOptions = {
  'hoodie.datasource.query.type': 'incremental',
  'hoodie.datasource.read.begin.instanttime': <beginInstantTime>,
}

incQueryDF = spark.read \
    .format('org.apache.hudi') \
    .options(**readOptions) \
    .load('s3://DOC-EXAMPLE-BUCKET/myhudidataset')

incQueryDF.show()
```

For more information about reading from Hudi datasets, see Querying Hudi tables in the Apache Hudi documentation.

Use the Hudi CLI

You can use the Hudi CLI to administer Hudi datasets to view information about commits, the filesystem, statistics, and more. You can also use the CLI to manually perform compactions, schedule compactions, or cancel scheduled compactions. For more information, see Interacting via CLI in the Apache Hudi documentation.

To start the Hudi CLI and connect to a dataset

1. Connect to the master node using SSH. For more information, see Connect to the master node using SSH in the *Amazon EMR Management Guide*.
2. At the command line, type `/usr/lib/hudi/cli/bin/hudi-cli.sh`.

   The command prompt changes to `hudi->`.
3. Type the following code to connect to a dataset. Replace *`s3://DOC-EXAMPLE-BUCKET/myhudidataset`* with the path to the dataset that you want to work with. The value we use is the same as the value established in earlier examples.

```
connect --path s3://DOC-EXAMPLE-BUCKET/myhudidataset
```

The command prompt changes to include the dataset that you're connected to, as shown in the following example.

```
hudi:myhudidataset->
```

Hudi release history

The following table lists the version of Hudi included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in Amazon EMR 7.x release versions, Amazon EMR 6.x release versions, or Amazon EMR 5.x release versions.

Hudi version information

Amazon EMR Release label	Hudi Version	Components installed with Hudi
emr-7.2.0	0.14.1-amzn-1	Not available.
emr-5.36.2	0.10.1-amzn-1	Not available.
emr-7.1.0	0.14.1-amzn-0	Not available.
emr-7.0.0	0.14.0-amzn-1	Not available.
emr-6.15.0	0.14.0-amzn-0	Not available.
emr-6.14.0	0.13.1-amzn-2	Not available.
emr-6.13.0	0.13.1-amzn-1	Not available.
emr-6.12.0	0.13.1-amzn-0	Not available.
emr-6.11.1	0.13.0-amzn-0	Not available.
emr-6.11.0	0.13.0-amzn-0	Not available.

Amazon EMR Release label	Hudi Version	Components installed with Hudi
emr-6.10.1	0.12.2-amzn-0	Not available.
emr-6.10.0	0.12.2-amzn-0	Not available.
emr-6.9.1	0.12.1-amzn-0	Not available.
emr-6.9.0	0.12.1-amzn-0	Not available.
emr-6.8.1	0.11.1-amzn-0	Not available.
emr-6.8.0	0.11.1-amzn-0	Not available.
emr-6.7.0	0.11.0-amzn-0	Not available.
emr-5.36.1	0.10.1-amzn-1	Not available.
emr-5.36.0	0.10.1-amzn-1	Not available.
emr-6.6.0	0.10.1-amzn-0	Not available.
emr-5.35.0	0.9.0-amzn-2	Not available.
emr-6.5.0	0.9.0-amzn-1	Not available.
emr-6.4.0	0.8.0-amzn-0	Not available.
emr-6.3.1	0.7.0-amzn-0	Not available.
emr-6.3.0	0.7.0-amzn-0	Not available.
emr-6.2.1	0.6.0-amzn-1	Not available.
emr-6.2.0	0.6.0-amzn-1	Not available.
emr-6.1.1	0.5.2-incubating-amzn-2	Not available.
emr-6.1.0	0.5.2-incubating-amzn-2	Not available.
emr-6.0.1	0.5.0-incubating-amzn-1	Not available.

Amazon EMR Release label	Hudi Version	Components installed with Hudi
emr-6.0.0	0.5.0-incubating-amzn-1	Not available.
emr-5.34.0	0.9.0-amzn-0	Not available.
emr-5.33.1	0.7.0-amzn-1	Not available.
emr-5.33.0	0.7.0-amzn-1	Not available.
emr-5.32.1	0.6.0-amzn-0	Not available.
emr-5.32.0	0.6.0-amzn-0	Not available.
emr-5.31.1	0.6.0-amzn-0	Not available.
emr-5.31.0	0.6.0-amzn-0	Not available.
emr-5.30.2	0.5.2-incubating	Not available.
emr-5.30.1	0.5.2-incubating	Not available.
emr-5.30.0	0.5.2-incubating	Not available.
emr-5.29.0	0.5.0-incubating	Not available.
emr-5.28.1	0.5.0-incubating	Not available.
emr-5.28.0	0.5.0-incubating	Not available.

Hue

Hue (Hadoop User Experience) is an open-source, web-based, graphical user interface for use with Amazon EMR and Apache Hadoop. Hue groups together several different Hadoop ecosystem projects into a configurable interface. Amazon EMR has also added customizations specific to Hue in Amazon EMR. Hue acts as a front-end for applications that run on your cluster, allowing you to interact with applications using an interface that may be more familiar or user-friendly. The applications in Hue, such as the Hive and Pig editors, replace the need to log in to the cluster to run scripts interactively using each application's respective shell. After a cluster launches, you might interact entirely with applications using Hue or a similar interface. For more information about Hue, see http://gethue.com.

Hue is installed by default when you launch your cluster using the Amazon EMR console. You can choose not to install Hue by using **Advanced options** in the Amazon EMR console when you launch a cluster, or by explicitly specifying the `--applications` option, and omitting Hue, when you use `create-cluster` from the AWS CLI.

Topics

Hue version information

Hue version for 7.2.0

The following table lists the version of Hue included in the latest release of the Amazon EMR 7.x series, along with the components that Amazon EMR installs with Hue.

For the version of components installed with Hue in this release, see Release 7.2.0 Component Versions.

Hue version information for emr-7.2.0

Amazon EMR Release Label	Hue Version	Components Installed With Hue
emr-7.2.0	Hue 4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Hue version for 6.15.0

The following table lists the version of Hue included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with Hue.

For the version of components installed with Hue in this release, see Release 6.15.0 Component Versions.

Hue version information for emr-6.15.0

Amazon EMR Release Label	Hue Version	Components Installed With Hue
emr-6.15.0	Hue 4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-ya

Amazon EMR Release Label	Hue Version	Components Installed With Hue
		rn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Hue version for 5.36.2

The following table lists the version of Hue included in the latest release of the Amazon EMR 5.x series, along with the components that Amazon EMR installs with Hue.

For the version of components installed with Hue in this release, see Release 5.36.2 Component Versions.

Hue version information for emr-5.36.2

Amazon EMR Release Label	Hue Version	Components Installed With Hue
emr-5.36.2	Hue 4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Supported and unsupported features of Hue on Amazon EMR

With Amazon EMR 7.0.0 and higher, Hue requires Python 3.9 or higher. When you use in-transit encryption, you must use a 2048-bit cipher key.

- Amazon S3 and Hadoop File System (HDFS) Browser

  - With the appropriate permissions, you can browse and move data between the ephemeral HDFS storage and S3 buckets belonging to your account.
  - By default, superusers in Hue can access all files that Amazon EMR IAM roles are allowed to access. Newly created users do not automatically have permissions to access the Amazon S3 filebrowser and must have the `filebrowser.s3_access` permissions enabled for their group.
- Hive — Run interactive queries on your data. This is also a useful way to prototype programmatic or batched querying.
- Pig — Run scripts on your data or issue interactive commands.
- Oozie — Create and monitor Oozie workflows.
- Metastore Manager — View and manipulate the contents of the Hive metastore (import/create, drop, and so on).
- Job browser — See the status of your submitted Hadoop jobs.
- User management — Manage Hue accounts and integrate LDAP users with Hue.
- AWS Samples — There are several "ready-to-run" examples that process sample data from various AWS services using applications in Hue. When you log in to Hue, you are taken to the Hue Home application where the samples are pre-installed.
- Livy Server is supported only in Amazon EMR version 5.9.0 and later.
- To use the Hue Notebook for Spark, you must install Hue with Livy and Spark.
- Trino/Presto — Hue supports connecting to Trino and Presto data servers. With Amazon EMR 7.0.0 and higher, this connectivity is facilitated with `trino-python-client` and `presto-python-client` connectors. Releases lower than 7.0.0 use a JDBC connector through a JDBC gateway.
- The Hue Dashboard is not supported.
- PostgreSQL is not supported.

Considerations

Consider the following limitations when you use Hue on Amazon EMR.

Performance with large Hue metadata tables

If the Hue metadata database gets too big, performance might degrade. To check the size of the tables, first connect to the primary node of the Amazon EMR on Amazon EC2 cluster with SSH and run the command `sudo mysql -u root` to start the MySQL CLI. To get the size of your table, run the query `SELECT COUNT(*) FROM hue.<table_name>`. See the following for what *<table_name>* can be:

- desktop_document
- desktop_document2
- oozie_job
- beeswax_savedquery
- beeswax_session
- beeswax_queryhistory

If running that query returns a count of more than 100000, you should run the following clean up command to delete the old records.

```
cd /opt/cloudera/parcels/CDH/lib/hue # Hue home directory ./build/env/bin/hue
 desktop_document_cleanup
```

For more information about cleaning up your database, see the [reference page in the Hue documentation](#).

Hue doesn't automatically clean the tables, but Amazon EMR releases 5.12.0 and higher provide a method to delete old documents in the tables. Create the following shell script and run it as a step in an Amazon EMR cluster with an integer parameter that represents the maximum number of days of how long to keep documents in the metadata database.

```
#!/bin/bash
if grep isMaster /mnt/var/lib/info/instance.json | grep false;
then
  echo "This is not the primary node; do nothing, exiting"
  exit 0
fi
while [ ! -f /usr/lib/hue/desktop/core/src/desktop/management/commands/
desktop_document_cleanup.py ]
do
  sleep 1
done
sudo systemctl stop hue.service
sudo sed -i 's+  LOG.warn+  # LOG.warn+g' /usr/lib/hue/desktop/core/src/desktop/
management/commands/desktop_document_cleanup.py
```

```
sudo /usr/lib/hue/build/env/bin/hue desktop_document_cleanup --keep-days $1
sudo systemctl start hue.service
```

Incompatibility between Hue versions

If you're using the same Hue metadata database across multiple Hue-enabled clusters, we recommend that these clusters run the same version of Hue. Different versions of Hue can have different schemas for the Hue metadata database. Using the same database for different versions can cause a Hue installation to fail. For example, trying to use the same database for two clusters with 4.10.0 and 4.11.0 installed can cause login errors for the users trying to log in to the 4.10.0 cluster.

For a list of Amazon EMR release labels and the corresponding installed versions of Hue, see Hue release history.

Connecting to the Hue web user interface

Connecting to the Hue web user interface is the same as connecting to any HTTP interface hosted on the master node of a cluster. The following procedure describes how to access the Hue user interface. For more information, see View web interfaces hosted on EMR clusters in the *Amazon EMR Management Guide*.

To view the Hue web user interface

1. Follow these instructions to Set up an SSH tunnel to the master node using dynamic port forwarding in the *Amazon EMR Management Guide*.
2. Type the following address in your browser to open the **Hue** web interface: `http://master public DNS:8888` where *`master public dns`* is the public DNS name of your cluster master node, for example `ec2-11-22-333-44.compute-1.amazonaws.com`.
3. At the Hue login screen, if you are the administrator logging in for the first time, enter a user name and password to create your Hue superaccount and then select **Create account**. Otherwise, type your username and password and select **Create account** or enter the credentials provided by your administrator.

Using Hue with a remote database in Amazon RDS

By default, Hue user information and query histories are stored in a local MySQL database on the master node. Alternatively, you can create one or more Hue-enabled clusters using a configuration stored in Amazon S3 and a MySQL database in Amazon Relational Database Service(Amazon RDS). This allows you to persist user information and query history created by Hue without keeping your Amazon EMR cluster running. We recommend using Amazon S3 server-side encryption to store the configuration file.

First create the remote database for Hue.

To create the external MySQL database

1. Open the Amazon RDS console at https://console.aws.amazon.com/rds/.
2. Click **Launch a DB Instance**.
3. Choose MySQL and click **Select**.
4. Leave the default selection of **Multi-AZ Deployment and Provisioned IOPS Storage** and click **Next**.
5. Leave the Instance Specifications at their defaults, specify Settings, and click **Next**.
6. On the Configure Advanced Settings page, choose a proper security group and database name. The security group you use must at least allow ingress TCP access for port 3306 from the master node of your cluster. If you have not created your cluster at this point, you can allow all hosts to connect to port 3306 and adjust the security group after you have launched the cluster. Click **Launch DB Instance**.
7. From the RDS Dashboard, select **Instances** and select the instance you have just created. When your database is available, make a note of the dbname, username, password, and RDS instance hostname. You use this information when you create and configure your cluster.

To specify an external MySQL database for Hue when launching a cluster using the AWS CLI

To specify an external MySQL database for Hue when launching a cluster using the AWS CLI, use the information you noted when creating your RDS instance for configuring `hue.ini` with a configuration object

Note

You can create multiple clusters that use the same external database, but each cluster will share query history and user information.

- Using the AWS CLI, create a cluster with Hue installed, using the external database you created, and referencing a configuration file with a configuration classification for Hue that specifies the database properties. The following example creates a cluster with Hue installed, referencing a configuration file in Amazon S3, `myConfig.json`, that specifies the database configuration.

  Note

  Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --release-label emr-7.2.0 --applications Name=Hue Name=Spark
 Name=Hive \
--instance-type m5.xlarge --instance-count 3 \
--configurations https://s3.amazonaws.com/mybucket/myfolder/myConfig.json --use-
default-roles
```

  Example contents of the `myConfig.json` file are shown below. Replace *`dbname`*, *`username`*, *`password`*, and *`RDS instance hostname`* with the values that you noted earlier in the RDS Dashboard.

```
[{
  "Classification": "hue-ini",
  "Properties": {},
  "Configurations": [
    {
      "Classification": "desktop",
      "Properties": {},
      "Configurations": [
        {
          "Classification": "database",
          "Properties": {
```

```
            "name": "dbname",
            "user": "username",
            "password": "password",
            "host": "RDS instance hostname",
            "port": "3306",
            "engine": "mysql"
          },
          "Configurations": []
        }
      ]
    }
  ]
}]
```

Troubleshooting

In the event of Amazon RDS failover

It is possible users may encounter delays when running a query because the Hue database instance is non-responsive or is in the process of failover. The following are some facts and guidelines for this issue:

- If you login to the Amazon RDS console, you can search for failover events. For example, to see if a failover is in process or has occurred, look for events such as "Multi-AZ instance failover started" and "Multi-AZ instance failover completed."
- It takes about 30 seconds for an RDS instance to complete a failover.
- If you are experiencing longer-than-normal responses for queries in Hue, try to re-execute the query.

Advanced configurations for Hue

This section includes the following topics.

Topics

- Configure Hue for LDAP users

Configure Hue for LDAP users

Integration with LDAP allows users to log into Hue using existing credentials stored in an LDAP directory. When you integrate Hue with LDAP, you do not need to independently manage user information in Hue. The information below demonstrates Hue integration with Microsoft Active Directory, but the configuration options are analogous to any LDAP directory.

LDAP authentication first binds to the server and establishes the connection. Then, the established connection is used for any subsequent queries to search for LDAP user information. Unless your Active Directory server allows anonymous connections, a connection needs to be established using a bind distinguished name and password. The bind distinguished name (or DN) is defined by the `bind_dn` configuration setting. The bind password is defined by the `bind_password` configuration setting. Hue has two ways to bind LDAP requests: search bind and direct bind. The preferred method for using Hue with Amazon EMR is search bind.

When search bind is used with Active Directory, Hue uses the user name attribute (defined by `user_name_attr config`) to find the attribute that needs to be retrieved from the base distinguished name (or DN). Search bind is useful when the full DN is not known for the Hue user.

For example, you may have `user_name_attr config` set to use the common name (or CN). In that case, the Active Directory server uses the Hue username provided during login to search the directory tree for a common name that matches, starting at the base distinguished name. If the common name for the Hue user is found, the user's distinguished name is returned by the server. Hue then constructs a distinguished name used to authenticate the user by performing a bind operation.

> **Note**
>
> Search bind searches usernames in all directory subtrees beginning at the base distinguished name. The base distinguished name specified in the Hue LDAP configuration should be the closest parent of the username, or your LDAP authentication performance may suffer.

When direct bind is used with Active Directory, the exact `nt_domain` or `ldap_username_pattern` must be used to authenticate. When direct bind is used, if the nt domain (defined by the `nt_domain` configuration setting) attribute is defined, a user distinguished name template is created using the form: `<login username>@nt_domain`. This template is used to search all directory subtrees beginning at the base distinguished name. If the nt domain is not configured, Hue searches for an exact distinguished name pattern for the user (defined by the ldap_username_pattern configuration setting). In this instance, the server searches for a matching ldap_username_pattern value in all directory subtrees beginning at the base distinguished name.

To launch a cluster with LDAP properties for Hue using the AWS CLI

- To specify LDAP properties for hue-ini, create a cluster with Hue installed and reference a json file with configuration properties for LDAP. An example command is shown below, which references a configuration file myConfig.json stored in Amazon S3.

```
aws emr create-cluster --release-label emr-7.2.0 --applications Name=Hue Name=Spark
 Name=Hive \
--instance-type m5.xlarge --instance-count 3 --configurations https://
s3.amazonaws.com/mybucket/myfolder/myConfig.json.
```

Example contents of myConfig.json are shown below.

```
[
    {
        "Classification": "hue-ini",
        "Properties": {},
        "Configurations": [
            {
                "Classification": "desktop",
                "Properties": {},
                "Configurations": [
                    {
                        "Classification": "ldap",
                        "Properties": {},
                        "Configurations": [
                            {
                                "Classification": "ldap_servers",
                                "Properties": {},
                                "Configurations": [
                                    {
                                        "Classification": "yourcompany",
                                        "Properties": {
                                            "base_dn":
 "DC=yourcompany,DC=hue,DC=com",
                                            "ldap_url": "ldap://ldapurl",
                                            "search_bind_authentication": "true",
```

```
                                            "bind_dn":
 "CN=hue,CN=users,DC=yourcompany,DC=hue,DC=com",
                                            "bind_password": "password"
                                        },
                                        "Configurations": []
                                    }
                                ]
                            }
                        ]
                    },
                    {
                        "Classification": "auth",
                        "Properties": {
                         "backend": "desktop.auth.backend.LdapBackend"
                              }
                    }
                ]
            }

        ]
    }
]
```

> **Note**
>
> With Amazon EMR version 5.21.0 and later, you can override cluster configurations and specify additional configuration classifications for each instance group in a running cluster. You do this by using the Amazon EMR console, the AWS Command Line Interface (AWS CLI), or the AWS SDK. For more information, see Supplying a Configuration for an Instance Group in a Running Cluster.

To view LDAP settings in Hue

1. Verify you have an active VPN connection or SSH tunnel to the Amazon EMR cluster's master node. Then, in your browser, type *master-public-dns*:8888 to open the Hue web interface.
2. Log in using your Hue administrator credentials. If the **Did you know?** window opens, click **Got it, prof!** to close it.
3. Click the **Hue** icon in the toolbar.

4. On the **About Hue** page, select **Configuration**.
5. In the **Configuration Sections and Variables** section, click **Desktop**.
6. Scroll to the **ldap** section to view your settings.

Hue release history

The following table lists the version of Hue included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in Amazon EMR 7.x release versions, Amazon EMR 6.x release versions, or Amazon EMR 5.x release versions.

Hue version information

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-7.2.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.36.2	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-serve

Amazon EMR Release label	Hue Version	Components installed with Hue
		r, mariadb-server, oozie-client, oozie-server
emr-7.1.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-7.0.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-6.15.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.14.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-6.13.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.12.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-6.11.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.11.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-6.10.1	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.10.0	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-6.9.1	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.9.0	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-6.8.1	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.8.0	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-6.7.0	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.36.1	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.36.0	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.6.0	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.35.0	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.5.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-6.4.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.3.1	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-6.3.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.2.1	4.8.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-6.2.0	4.8.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.1.1	4.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-6.1.0	4.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.0.1	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-6.0.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.34.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.33.1	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.33.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.32.1	4.8.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.32.0	4.8.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.31.1	4.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.31.0	4.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.30.2	4.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.30.1	4.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.30.0	4.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.29.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.28.1	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.28.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.27.1	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.27.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.26.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.25.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.24.1	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.24.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.23.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.23.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.22.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.21.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.21.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.21.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.20.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.20.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.19.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.19.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.18.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.18.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.17.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.17.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.17.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.16.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.16.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.15.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.15.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.14.2	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.14.1	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.14.0	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.13.1	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.13.0	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.12.3	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.12.2	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.12.1	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.12.0	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.11.4	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.11.3	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.11.2	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.11.1	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.11.0	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.10.1	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.10.0	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.9.1	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.9.0	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.8.3	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.8.2	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.8.1	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.8.0	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.7.1	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.7.0	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.6.1	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.6.0	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.5.4	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.5.3	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.5.2	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.5.1	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.5.0	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.4.1	3.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.4.0	3.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.3.2	3.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.3.1	3.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.3.0	3.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.2.3	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.2.2	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.2.1	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.2.0	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.1.1	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.1.0	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.0.3	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-5.0.2	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-5.0.1	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-5.0.0	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-4.9.6	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.9.5	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.9.4	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-4.9.3	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.9.2	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.9.1	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-4.8.5	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.8.4	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.8.3	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-4.8.2	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.8.1	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.8.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-4.7.4	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.7.3	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.7.2	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-4.7.1	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.7.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.6.1	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-4.6.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.5.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.4.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server

Amazon EMR Release label	Hue Version	Components installed with Hue
emr-4.3.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.2.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.1.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server

Iceberg

Apache Iceberg is an open table format for large data sets in Amazon Simple Storage Service (Amazon S3). It provides fast query performance over large tables, atomic commits, concurrent writes, and SQL-compatible table evolution. Starting with Amazon EMR 6.5.0, you can use Apache Spark 3 on Amazon EMR clusters with the Iceberg table format.

The following table lists the version of Iceberg included in the latest release of the Amazon EMR 7.x series, along with the components that Amazon EMR installs with Iceberg.

For the version of components installed with Iceberg in this release, see Release 7.2.0 Component Versions.

Iceberg version information for emr-7.2.0

Amazon EMR Release Label	Iceberg Version	Components Installed With Iceberg
emr-7.2.0	Iceberg 1.5.0-amzn-0	Not available.

The following table lists the version of Iceberg included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with Iceberg.

For the version of components installed with Iceberg in this release, see Release 6.15.0 Component Versions.

Iceberg version information for emr-6.15.0

Amazon EMR Release Label	Iceberg Version	Components Installed With Iceberg
emr-6.15.0	Iceberg 1.4.0-amzn-0	Not available.

Topics

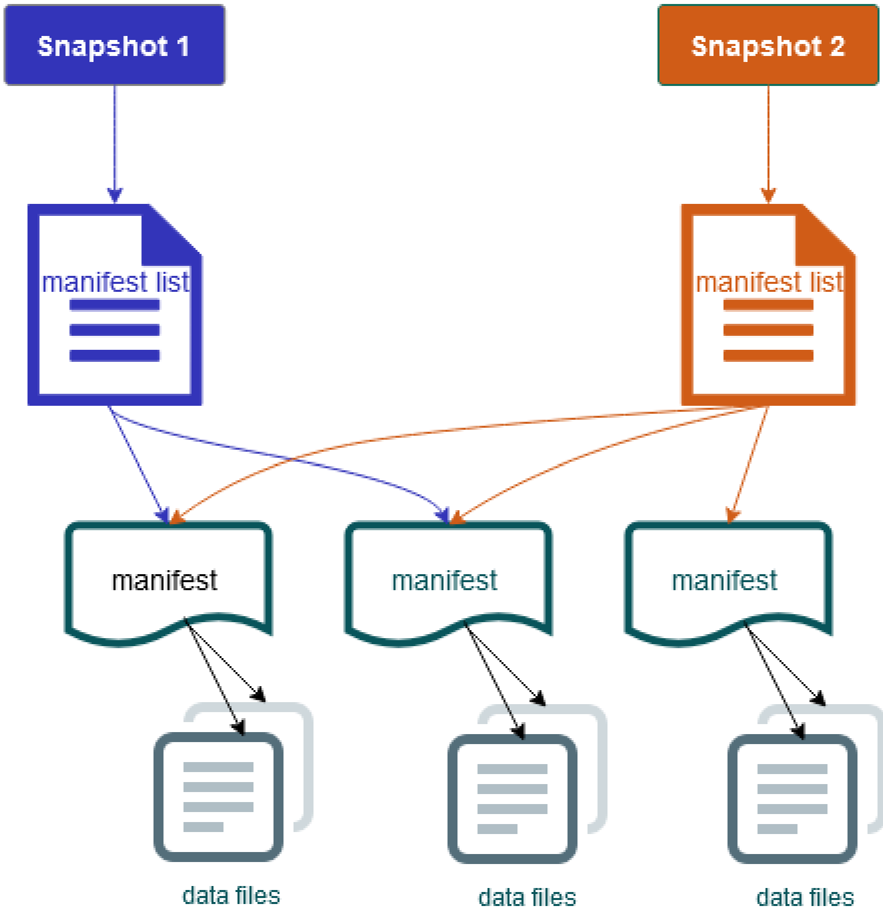
- [Considerations and limitations for using Iceberg on Amazon EMR]
- [Iceberg release history]

How Iceberg works

Iceberg tracks individual data files in a table instead of in directories. This way, writers can create data files in place (files are not moved or changed). Also, writers can only add files to the table in an explicit commit. The table state is maintained in metadata files. All changes to the table state create a new metadata file that atomically replaces the older metadata. The table metadata file tracks the table schema, partitioning configuration, and other properties.

It also includes snapshots of the table contents. Each snapshot is a complete set of data files in the table at a point in time. Snapshots are listed in the metadata file, but the files in a snapshot are stored in separate manifest files. The atomic transitions from one table metadata file to the next provide snapshot isolation. Readers use the snapshot that was current when they loaded the table metadata. Readers aren't affected by changes until they refresh and pick up a new metadata location. Data files in snapshots are stored in one or more manifest files that contain a row for each data file in the table, its partition data, and its metrics. A snapshot is the union of all files in its manifests. Manifest files can also be shared between snapshots to avoid rewriting metadata that changes infrequently.

Iceberg snapshot diagram



Iceberg offers the following features:

- Supports ACID transactions and time travel in your Amazon S3 data lake.
- Commit retries benefit from the performance advantages of optimistic concurrency.
- File-level conflict resolution results in high concurrency.
- With min-max statistics per column in metadata, you can skip files, which boosts performance for selective queries.
- You can organize tables into flexible partition layouts, with partition evolution enabling updates to partition schemes. Queries and data volumes can then change without relying on physical directories.
- Supports schema evolution and enforcement.
- Iceberg tables act as idempotent sinks and replayable sources. This enables streaming and batch support with exactly-once pipelines. Idempotent sinks track write operations that have succeeded in the past. Therefore, the sink can request data again in case of a failure, and drop data if it has been sent multiple times.

- View history and lineage, including table evolution, operations history, and statistics for each commit.
- Migrate from an existing dataset with a choice of data format (Parquet, ORC, Avro) and analytics engine (Spark, Trino, PrestoDB, Flink, Hive).

Use a cluster with Iceberg installed

This section includes information for using Iceberg with Spark, Trino, Flink, and Hive.

Use an Iceberg cluster with Spark

Starting with Amazon EMR version 6.5.0, you can use Iceberg with your Spark cluster with no requirement to include bootstrap actions. For Amazon EMR versions 6.4.0 and earlier, you can use a bootstrap action to pre-install all necessary dependencies.

In this tutorial, you use the AWS CLI to work with Iceberg on an Amazon EMR Spark cluster. To use the console to create a cluster with Iceberg installed, follow the steps in Build an Apache Iceberg data lake using Amazon Athena, Amazon EMR, and AWS Glue.

Create an Iceberg cluster

You can create a cluster with Iceberg installed using the AWS Management Console, the AWS CLI or the Amazon EMR API. In this tutorial, you use the AWS CLI to work with Iceberg on an Amazon EMR cluster. To use the console to create a cluster with Iceberg installed, follow the steps in Build an Apache Iceberg data lake using Amazon Athena, Amazon EMR, and AWS Glue.

To use Iceberg on Amazon EMR with the AWS CLI, first create a cluster with the following steps. For information on specifying the Iceberg classification using the AWS CLI, see Supply a configuration using the AWS CLI when you create a cluster or Supply a configuration using the Java SDK when you create a cluster.

1. Create a `configurations.json` file with the following content:

```
[{
    "Classification":"iceberg-defaults",
    "Properties":{"iceberg.enabled":"true"}
}]
```

2. Next, create a cluster with the following configuration. Replace the example Amazon S3 bucket path and the subnet ID with your own.

```
aws emr create-cluster --release-label emr-6.5.0 \
--applications Name=Spark \
--configurations file://iceberg_configurations.json \
--region us-east-1 \
--name My_Spark_Iceberg_Cluster \
--log-uri s3://DOC-EXAMPLE-BUCKET/ \
--instance-type m5.xlarge \
--instance-count 2 \
--service-role EMR_DefaultRole_V2 \
--ec2-attributes
 InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef0
```

Alternatively, you can create an Amazon EMR cluster including the Spark application and include the file `/usr/share/aws/iceberg/lib/iceberg-spark3-runtime.jar` as a JAR dependency in a Spark job. For more information, see Submitting Applications.

To include the jar as a dependency in a Spark job, add the following configuration property to the Spark application:

```
--conf "spark.jars=/usr/share/aws/iceberg/lib/iceberg-spark3-runtime.jar"
```

For more information about Spark job dependencies, see Dependency Management in the Apache Spark document Running Spark on Kubernetes.

Initialize a Spark session for Iceberg

The following examples demonstrate how to launch the interactive Spark shell, use Spark submit, or use Amazon EMR Notebooks to work with Iceberg on Amazon EMR.

spark-shell

1. Connect to the master node using SSH. For more information, see Connect to the master node using SSH in the *Amazon EMR Management Guide*.
2. Enter the following command to launch the Spark shell. To use the PySpark shell, replace `spark-shell` with `pyspark`.

```
spark-shell \
```

```
--conf
 "spark.sql.extensions=org.apache.iceberg.spark.extensions.IcebergSparkSessionExtensions"
 \
--conf "spark.sql.catalog.dev=org.apache.iceberg.spark.SparkCatalog" \
--conf "spark.sql.catalog.dev.type=hadoop" \
--conf "spark.sql.catalog.dev.warehouse=s3://DOC-EXAMPLE-BUCKET/example-prefix/"
```

spark-submit

1. Connect to the master node using SSH. For more information, see Connect to the master node using SSH in the *Amazon EMR Management Guide*.
2. Enter the following command to launch the Spark session for Iceberg.

```
spark-submit \
--conf
 "spark.sql.extensions=org.apache.iceberg.spark.extensions.IcebergSparkSessionExtensions"
 \
--conf "spark.sql.catalog.dev=org.apache.iceberg.spark.SparkCatalog" \
--conf "spark.sql.catalog.dev.type=hadoop" \
--conf "spark.sql.catalog.dev.warehouse=s3://DOC-EXAMPLE-BUCKET/example-prefix/"
```

EMR Studio notebooks

To initialize a Spark session using EMR Studio notebooks, configure your Spark session using the `%%configure` magic command in your Amazon EMR notebook, as in the following example. For more information, see Use EMR Notebooks magics in the *Amazon EMR Management Guide*.

```
%%configure -f
{
"conf":{

 "spark.sql.extensions":"org.apache.iceberg.spark.extensions.IcebergSparkSessionExtensions",
    "spark.sql.catalog.dev":"org.apache.iceberg.spark.SparkCatalog",
    "spark.sql.catalog.dev.type":"hadoop",
    "spark.sql.catalog.dev.warehouse":"s3://DOC-EXAMPLE-BUCKET/example-prefix/"
    }
}
```

Write to an Iceberg table

The following example shows how to create a DataFrame and write it as an Iceberg dataset. The examples demonstrate working with datasets using the Spark shell while connected to the master node using SSH as the default hadoop user.

> **Note**
>
> To paste code samples into the Spark shell, type `:paste` at the prompt, paste the example, and then press CTRL+D.

PySpark

Spark includes a Python-based shell, `pyspark`, that you can use to prototype Spark programs written in Python. Invoke `pyspark` on the master node.

```
## Create a DataFrame.
data = spark.createDataFrame([
 ("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
 ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
 ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
 ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z")
],["id", "creation_date", "last_update_time"])

## Write a DataFrame as a Iceberg dataset to the Amazon S3 location.
spark.sql("""CREATE TABLE IF NOT EXISTS dev.db.iceberg_table (id string,
creation_date string,
last_update_time string)
USING iceberg
location 's3://DOC-EXAMPLE-BUCKET/example-prefix/db/iceberg_table'""")

data.writeTo("dev.db.iceberg_table").append()
```

Scala

```
import org.apache.spark.sql.SaveMode
import org.apache.spark.sql.functions._

// Create a DataFrame.
val data = Seq(
```

```
("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
("103", "2015-01-01", "2015-01-01T13:51:40.519832Z")
).toDF("id", "creation_date", "last_update_time")

// Write a DataFrame as a Iceberg dataset to the Amazon S3 location.
spark.sql("""CREATE TABLE IF NOT EXISTS dev.db.iceberg_table (id string,
creation_date string,
last_update_time string)
USING iceberg
location 's3://DOC-EXAMPLE-BUCKET/example-prefix/db/iceberg_table'""")

data.writeTo("dev.db.iceberg_table").append()
```

Read from an Iceberg table

PySpark

```
df = spark.read.format("iceberg").load("dev.db.iceberg_table")
df.show()
```

Scala

```
val df = spark.read.format("iceberg").load("dev.db.iceberg_table")
df.show()
```

Spark SQL

```
SELECT * from dev.db.iceberg_table LIMIT 10
```

Configure Spark properties to use the AWS Glue Data Catalog as Iceberg tables metastore

To use the AWS Glue Catalog as the Metastore for Iceberg tables, set the Spark configuration properties as below:

```
spark-submit \
```

```
    --conf spark.sql.catalog.my_catalog=org.apache.iceberg.spark.SparkCatalog \
    --conf spark.sql.catalog.my_catalog.warehouse=s3://<bucket>/<prefix> \
    --conf spark.sql.catalog.my_catalog.catalog-
impl=org.apache.iceberg.aws.glue.GlueCatalog \
    --conf spark.sql.catalog.my_catalog.io-impl=org.apache.iceberg.aws.s3.S3FileIO \
    --conf spark.sql.catalog.my_catalog.lock-
impl=org.apache.iceberg.aws.dynamodb.DynamoDbLockManager \
    --conf spark.sql.catalog.my_catalog.lock.table=myGlueLockTable
```

Use an Iceberg cluster with Trino

Starting with Amazon EMR version 6.6.0, you can use Iceberg with your Trino cluster.

In this tutorial, you use the AWS CLI to work with Iceberg on an Amazon EMR Trino cluster. To use the console to create a cluster with Iceberg installed, follow the steps in Build an Apache Iceberg data lake using Amazon Athena, Amazon EMR, and AWS Glue.

Create an Iceberg cluster

To use Iceberg on Amazon EMR with the AWS CLI, first create a cluster with the following steps. For information on specifying the Iceberg classification using the AWS CLI, see Supply a configuration using the AWS CLI when you create a cluster or Supply a configuration using the Java SDK when you create a cluster.

1. Create an `iceberg.properties` file and set a value for your chosen catalog. For example, if you want to use the Hive metastore as your catalog, your file should have the following content.

   ```
   connector.name=iceberg
   hive.metastore.uri=thrift://localhost:9083
   ```

   If you want to use the AWS Glue Data Catalog as your store, your file should have the following content.

   ```
   connector.name=iceberg
   iceberg.catalog.type=glue
   ```

2. Create a bootstrap action that copies `iceberg.properties` from Amazon S3 to `/etc/trino/conf/catalog/iceberg.properties`, as in the following example. For information on bootstrap actions, see Create bootstrap actions to install additional software.

```
set -ex
sudo aws s3 cp s3://DOC-EXAMPLE-BUCKET/iceberg.properties /etc/trino/conf/catalog/
iceberg.properties
```

3. Create a cluster with the following configuration, replacing the example bootstrap actions script path and key name with your own.

```
aws emr create-cluster --release-label emr-6.7.0 \
--applications Name=Trino \
--region us-east-1 \
--name My_Trino_Iceberg_Cluster \
--bootstrap-actions '[{"Path":"s3://DOC-EXAMPLE-BUCKET","Name":"Add
 iceberg.properties"}]' \
--instance-groups InstanceGroupType=MASTER,InstanceCount=1,InstanceType=c3.4xlarge
 InstanceGroupType=CORE,InstanceCount=3,InstanceType=c3.4xlarge \
--use-default-roles \
--ec2-attributes KeyName=<key-name>
```

Initialize a Trino session for Iceberg

To initialize a Trino session, run the following command.

```
trino-cli --catalog iceberg
```

Write to an Iceberg table

Create and write to your table with the following SQL commands.

```
trino> SHOW SCHEMAS;
trino> CREATE TABLE default.iceberg_table (
            id int,
            data varchar,
            category varchar)
       WITH (
            format = 'PARQUET',
            partitioning = ARRAY['category', 'bucket(id, 16)'],
            location = 's3://DOC-EXAMPLE-BUCKET/<prefix>')

trino> INSERT INTO default.iceberg_table VALUES (1,'a','c1'), (2,'b','c2'),
 (3,'c','c3');
```

Read from a table for Iceberg

To read from your Iceberg table, run the following command.

```
trino> SELECT * from default.iceberg_table;
```

Use an Iceberg cluster with Flink

Starting with Amazon EMR version 6.9.0, you can use Iceberg with a Flink cluster without the setup steps required when using the open source Iceberg Flink Integration.

Creating an Iceberg cluster

You can create a cluster with Iceberg installed using the AWS Management Console, the AWS CLI, or the Amazon EMR API. In this tutorial, you use the AWS CLI to work with Iceberg on an Amazon EMR cluster. To use the console to create a cluster with Iceberg installed, follow the steps in Build an Apache Iceberg data lake using Amazon Athena, Amazon EMR, and AWS Glue.

To use Iceberg on Amazon EMR with the AWS CLI, first create a cluster with the following steps. For information on specifying the Iceberg classification using the AWS CLI, see Supply a configuration using the AWS CLI when you create a cluster or Supply a configuration using the Java SDK when you create a cluster. Create a file called `configurations.json` with the following content:

```
[{
"Classification":"iceberg-defaults",
    "Properties":{"iceberg.enabled":"true"}
}]
```

Next, create a cluster with the following configuration, replacing the example Amazon S3 bucket path and subnet ID with your own values:

```
aws emr create-cluster --release-label emr-6.9.0 \
--applications Name=Flink \
--configurations file://iceberg_configurations.json \
--region us-east-1 \
--name My_flink_Iceberg_Cluster \
--log-uri s3://DOC-EXAMPLE-BUCKET/ \
--instance-type m5.xlarge \
--instance-count 2 \
--service-role EMR_DefaultRole \
```

```
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef
```

Alternatively, you can create an Amazon EMR 6.9.0 cluster with a Flink application in it and use the file `/usr/share/aws/iceberg/lib/iceberg-flink-runtime.jar` as a JAR dependency in a Flink job.

Using the Flink SQL Client

The SQL Client script is located under `/usr/lib/flink/bin`. You can run the script with the following command:

```
flink-yarn-session -d # starting the Flink YARN Session in detached mode
./sql-client.sh
```

This launches a Flink SQL Shell.

Flink examples

Create an Iceberg table

Flink SQL

```
CREATE CATALOG glue_catalog WITH (
   'type'='iceberg',
   'warehouse'='<WAREHOUSE>',
   'catalog-impl'='org.apache.iceberg.aws.glue.GlueCatalog',
   'io-impl'='org.apache.iceberg.aws.s3.S3FileIO',
   'lock-impl'='org.apache.iceberg.aws.dynamodb.DynamoDbLockManager',
   'lock.table'='myGlueLockTable'
 );

USE CATALOG  glue_catalog;

CREATE DATABASE IF NOT EXISTS <DB>;

USE <DB>;

CREATE TABLE IF NOT EXISTS `glue_catalog`.`<DB>`.`sample` (id int, data string);
```

Table API

```
EnvironmentSettings settings =
```

```
                EnvironmentSettings.newInstance().inBatchMode().build();

TableEnvironment tEnv = TableEnvironment.create(settings);

String warehouse = "<WAREHOUSE>";
String db = "<DB>";

tEnv.executeSql(
                "CREATE CATALOG glue_catalog WITH (\n"
                        + "   'type'='iceberg',\n"
                        + "   'warehouse'='"
                        + warehouse
                        + "',\n"
                        + "   'catalog-impl'='org.apache.iceberg.aws.glue.GlueCatalog',
\n"
                        + "   'io-impl'='org.apache.iceberg.aws.s3.S3FileIO'\n"
                        + " );");

tEnv.executeSql("USE CATALOG  glue_catalog;");
tEnv.executeSql("CREATE DATABASE IF NOT EXISTS " + db + ";");
tEnv.executeSql("USE " + db + ";");
tEnv.executeSql(
        "CREATE TABLE `glue_catalog`.`" + db + "`.`sample` (id bigint, data string);");
```

Write to an Iceberg table

Flink SQL

```
INSERT INTO `glue_catalog`.`<DB>`.`sample` values (1, 'a'),(2,'b'),(3,'c');
```

Table API

```
tEnv.executeSql(
        "INSERT INTO `glue_catalog`.`"
                + db
                + "`.`sample` values (1, 'a'),(2,'b'),(3,'c');");
```

Datastream API

```
final StreamExecutionEnvironment env =
 StreamExecutionEnvironment.getExecutionEnvironment();
```

```
final StreamTableEnvironment tableEnv = StreamTableEnvironment.create(env);

String db = "<DB Name>";

String warehouse = "<Warehouse Path>";

GenericRowData rowData1 = new GenericRowData(2);
rowData1.setField(0, 1L);
rowData1.setField(1, StringData.fromString("a"));

DataStream<RowData> input = env.fromElements(rowData1);

Map<String, String> props = new HashMap<();
props.put("type", "iceberg");
props.put("warehouse", warehouse);
props.put("io-impl", "org.apache.iceberg.aws.s3.S3FileIO");

CatalogLoader glueCatlogLoader =
        CatalogLoader.custom(
                "glue",
                props,
                new Configuration(),
                "org.apache.iceberg.aws.glue.GlueCatalog");

TableLoader tableLoader =
        TableLoader.fromCatalog(glueCatlogLoader, TableIdentifier.of(db, "sample"));

DataStreamSink<Void> dataStreamSink =
        FlinkSink.forRowData(input).tableLoader(tableLoader).append();

env.execute("Datastream Write");
```

Read from an Iceberg table

Flink SQL

```
SELECT * FROM `glue_catalog`.`<DB>`.`sample`;
```

Table API

```
Table result = tEnv.sqlQuery("select * from `glue_catalog`.`" + db + "`.`sample`;");
```

Datastream API

```
final StreamExecutionEnvironment env =
 StreamExecutionEnvironment.getExecutionEnvironment();

final StreamTableEnvironment tableEnv = StreamTableEnvironment.create(env);

String db = "<DB Name>";

String warehouse = "<Warehouse Path>";

Map<String, String> props = new HashMap<>();
props.put("type", "iceberg");
props.put("warehouse", warehouse);
props.put("io-impl", "org.apache.iceberg.aws.s3.S3FileIO");

CatalogLoader glueCatlogLoader =
        CatalogLoader.custom(
                "glue",
                props,
                new Configuration(),
                "org.apache.iceberg.aws.glue.GlueCatalog");

TableLoader tableLoader =
        TableLoader.fromCatalog(glueCatlogLoader, TableIdentifier.of(db, "sample"));

DataStream<RowData> batch =

 FlinkSource.forRowData().env(env).tableLoader(tableLoader).streaming(false).build();

batch.print().name("print-sink");
```

Using the Hive catalog

Make sure the Flink and Hive dependencies are resolved as described in Configure Flink with Hive Metastore and Glue Catalog.

Running a Flink Job

One way to submit a job to Flink is to use a per job Flink YARN session. This can be launched with the following command:

```
sudo flink run -m yarn-cluster -p 4 -yjm 1024m -ytm 4096m $JAR_FILE_NAME
```

Use an Iceberg cluster with Hive

With Amazon EMR releases 6.9.0 and higher, you can use Iceberg with a Hive cluster without having to perform the setup steps that are required for Open Source Iceberg Hive Integration. For Amazon EMR versions 6.8.0 and earlier, you can use a bootstrap action to install `iceberg-hive-runtime` jar for configuring Hive for Iceberg support.

Amazon EMR 6.9.0 includes all features for Hive 3.1.3 integration with Iceberg 0.14.1 and also includes Amazon EMR added features such as auto selection of supported execution engines at runtime (Amazon EMR on EKS 6.9.0).

Create an Iceberg cluster

You can create a cluster with Iceberg installed using the AWS Management Console, the AWS CLI or the Amazon EMR API. In this tutorial, you use the AWS CLI to work with Iceberg on an Amazon EMR cluster. To use the console to create a cluster with Iceberg installed, follow the steps in Build an Iceberg data lake using Amazon Athena, Amazon EMR, and AWS Glue.

To use Iceberg on Amazon EMR with the AWS CLI, first create a cluster using the steps below. For information on specifying the Iceberg classification using the AWS CLI or the Java SDK, see Supply a configuration using the AWS CLI when you create a cluster or Supply a configuration using the Java SDK when you create a cluster. Create a file named `configurations.json` with the following content:

```
[{
    "Classification":"iceberg-defaults",
    "Properties":{"iceberg.enabled":"true"}
}]
```

Next, create a cluster with the following configuration, replacing the example Amazon S3 bucket path and the subnet ID with your own:

```
aws emr create-cluster --release-label emr-6.9.0 \
--applications Name=Hive \
--configurations file://iceberg_configurations.json \
--region us-east-1 \
--name My_hive_Iceberg_Cluster \
--log-uri s3://DOC-EXAMPLE-BUCKET/ \
--instance-type m5.xlarge \
--instance-count 2 \
```

```
--service-role EMR_DefaultRole \
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef
```

A Hive Iceberg cluster does the following things:

- Loads the Iceberg Hive runtime jar in Hive and enables Iceberg-related configuration for the Hive engine.
- Enables Amazon EMR Hive's dynamic execution engine selection to prevent users from setting supported execution engine for Iceberg compatibility.

> **Note**
>
> Hive Iceberg clusters do not currently support AWS Glue Data Catalog. The default Iceberg catalog is `HiveCatalog`, which corresponds to the metastore configured for the Hive environment. For more information on catalog management, see Using HCatalog in the Apache Hive documentation.

Feature support

Amazon EMR 6.9.0 supports Hive 3.1.3 and Iceberg 0.14.1. The feature support is limited to Iceberg-compatible features for Hive 3.1.2 and 3.1.3. The following commands are supported:

- With Amazon EMR releases 6.9.0 to 6.12.x, you must include the `libfb303` jar in the Hive `auxlib` directory. Use the following command to include it:

  ```
  sudo /usr/bin/ln -sf /usr/lib/hive/lib/libfb303-*.jar /usr/lib/hive/auxlib/
  libfb303.jar
  ```

  With Amazon EMR releases 6.13 and higher, the `libfb303` jar is automatically symlinked to the Hive `auxlib` directory.
- **Creating a table**
  - **Non-partitioned table** – External tables in Hive can be created by providing the storage handler as follows:

    ```
    CREATE EXTERNAL TABLE x (i int) STORED BY
     'org.apache.iceberg.mr.hive.HiveIcebergStorageHandler'
    ```

- **Partitioned table** – External partitioned tables in Hive can be created as follows:

```
CREATE EXTERNAL TABLE x (i int) PARTITIONED BY (j int) STORED BY
 'org.apache.iceberg.mr.hive.HiveIcebergStorageHandler'
```

> **Note**
>
> The `STORED AS` file format of ORC/AVRO/PARQUET is not supported in Hive 3. The default and only option is Parquet.

- **Dropping a table** – The `DROP TABLE` command is used to drop tables, such as in the following example:

```
DROP TABLE [IF EXISTS] table_name [PURGE];
```

- **Reading a table** – `SELECT` statements can be used to read Iceberg tables in Hive, such as in the following example. Supported execution engines are MR and Tez.

```
SELECT * FROM table_name
```

  For information on Hive's select syntax, see LanguageManual Select. For information on select statements with Iceberg tables in Hive, see Apache Iceberg Select.

- **Inserting into a table** – HiveQL's `INSERT INTO` statement works on Iceberg tables with support for the Map Reduce execution engine only. Amazon EMR users don't need to explicitly set the execution engine because Amazon EMR Hive selects the engine for Iceberg Tables at runtime.
  - **Single table insert into** – Example:

```
INSERT INTO table_name VALUES ('a', 1);
INSERT INTO table_name SELECT...;
```

  - **Multi-table insert into** – Non-atomic multi-table insert into statements are supported. Example:

```
FROM source
 INSERT INTO table_1 SELECT a, b
 INSERT INTO table_2 SELECT c,d;
```

Considerations and limitations for using Iceberg on Amazon EMR

This section includes considerations and limitations for using Iceberg with Spark, Trino, Flink, and Hive.

Considerations for using Iceberg with Spark

- Amazon EMR 6.5.0 does not support Iceberg running on Amazon EMR on EKS by default. An Amazon EMR 6.5.0 custom image is available so that you can pass `--jars local:///usr/share/aws/iceberg/lib/iceberg-spark3-runtime.jar` as a `spark-submit` parameter to create Iceberg tables on Amazon EMR on EKS. For more information, see Submit a Spark workload in Amazon EMR using a custom image in the *Amazon EMR on EKS Development Guide*. You can also contact AWS Support for assistance. Starting with Amazon EMR 6.6.0, Iceberg is supported on Amazon EMR on EKS.
- When using AWS Glue as a catalog for Iceberg, make sure the database in which you are creating a table exists in AWS Glue. If you are using services such as AWS Lake Formation and you're unable to load the catalog, make sure you have proper access to the service to execute the command.

Considerations for using Iceberg with Trino

- Amazon EMR 6.5 does not offer Trino Iceberg Catalog support for Iceberg natively. Trino needs Iceberg v0.11, so we recommend launching an Amazon EMR cluster for Trino separate from the Spark cluster and including Iceberg v0.11 on that cluster.
- When using AWS Glue as a catalog for Iceberg, make sure that the database in which you are creating a table exists in AWS Glue. If you are using services such as AWS Lake Formation and you're unable to load the catalog, make sure you have proper access to the service to execute the command.

Considerations for using Iceberg with Flink

When using AWS Glue as a catalog for Iceberg, make sure the database in which you are creating a table exists in AWS Glue. If you are using services such as AWS Lake Formation and you're unable to load the catalog, make sure you have proper access to the service to execute the command.

Considerations for using Iceberg with Hive

- Iceberg supports the following query types:
  - Create table
  - Drop table
  - Insert into table
  - Read table
- Only the MR (MapReduce) execution engine is supported for DML (data manipulation language) operations, and MR is deprecated in Hive 3.1.3.
- AWS Glue Data Catalog is not currently supported for Iceberg with Hive.
- Error handling is insufficiently robust. In cases of misconfiguration, inserts into queries might complete successfully. However, failure to update metadata can result in data loss.

Iceberg release history

The following table lists the version of Iceberg included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in Amazon EMR 7.x release versions, Amazon EMR 6.x release versions, or Amazon EMR 5.x release versions.

Iceberg version information

Amazon EMR Release label	Iceberg Version	Components installed with Iceberg
emr-7.2.0	1.5.0-amzn-0	Not available.
emr-7.1.0	1.4.3-amzn-0	Not available.
emr-7.0.0	1.4.2-amzn-0	Not available.
emr-6.15.0	1.4.0-amzn-0	Not available.

Amazon EMR Release label	Iceberg Version	Components installed with Iceberg
emr-6.14.0	1.3.1-amzn-0	Not available.
emr-6.13.0	1.3.0-amzn-1	Not available.
emr-6.12.0	1.3.0-amzn-0	Not available.
emr-6.11.1	1.2.0-amzn-0	Not available.
emr-6.11.0	1.2.0-amzn-0	Not available.
emr-6.10.1	1.1.0-amzn-0	Not available.
emr-6.10.0	1.1.0-amzn-0	Not available.
emr-6.9.1	0.14.1-amzn-0	Not available.
emr-6.9.0	0.14.1-amzn-0	Not available.
emr-6.8.1	0.14.0-amzn-0	Not available.
emr-6.8.0	0.14.0-amzn-0	Not available.
emr-6.7.0	0.13.1-amzn-0	Not available.
emr-6.6.0	0.13.1	Not available.
emr-6.5.0	0.12.0	Not available.

Iceberg release notes by version

- Amazon EMR 6.9.0 - Iceberg release notes

Amazon EMR 6.9.0 - Iceberg release notes

Amazon EMR 6.9.0 - Iceberg changes

Type	Description
Feature	Amazon EMR Flink integration with Iceberg.
Feature	Amazon EMR Hive integration with Iceberg.
Feature	Support to cache Iceberg metadata files on Amazon FSx for Lustre to improve the query planning time.
Backport	PR 5050: Flink 1.15: Support write options in the in-line insert SQL comments.
Backport	PR 5282: AWS: Fix PUT retry failures by opening new data file streams.
Backport	PR 5318: Flink 1.15: Bridge the gap between FlinkSource and IcebergSource (FLIP-27) and added an opt-in config to use FLIP-27 source in Flink SQL.
Backport	PR 5344: Flink 1.14: Bridge the gap between FlinkSource and IcebergSource (FLIP-27) and added an opt-in config to use FLIP-27 source in Flink SQL.
Backport	PR 5393: Flink 1.14, 1.15: Avoid convertin g Iceberg MetricContext to Flink metrics in FLIP-27 source reader.
Backport	PR 5401: Flink 1.14, 1.15: Missed IcebergSo urceReader group in PR #5393 for FLIP-27 source reader metrics.

Type	Description
Backport	PR 5679: Spark 3.2, 3.3: Fix nullability propagation for MergeRows node.
Backport	PR 5860: Spark 3.3: Fix QueryFailure when running RewriteManifestProcedure on Date partitioned tables.
Backport	PR 5880: Spark 3.3: Fix nullability in merge-on-read projections.
Backport	PR 5917: Spark 3.2: Fix nullability in merge-on-read projections.

Jupyter Notebook on Amazon EMR

Jupyter Notebook is an open-source web application that you can use to create and share documents that contain live code, equations, visualizations, and narrative text. Amazon EMR offers you three options to work with Jupyter notebooks:

Topics

EMR Studio

Amazon EMR Studio is a web-based integrated development environment (IDE) for fully managed Jupyter notebooks that run on Amazon EMR clusters. You can set up an EMR Studio for your team to develop, visualize, and debug applications written in R, Python, Scala, and PySpark.

We recommend using EMR Studio when using Jupyter notebooks on Amazon EMR. For more information, see EMR Studio in the *Amazon EMR Management Guide*.

Amazon EMR Notebook based on Jupyter Notebook

EMR Notebooks is a Jupyter Notebook environment built in to the Amazon EMR console that allows you to quickly create Jupyter notebooks, attach them to Spark clusters, and then open the Jupyter Notebook editor in the console to remotely run queries and code. An EMR notebook is saved in Amazon S3 independently from clusters for durable storage, quick access, and flexibility. You can have multiple notebooks open, attach multiple notebooks to a single cluster, and re-use a notebook on different clusters.

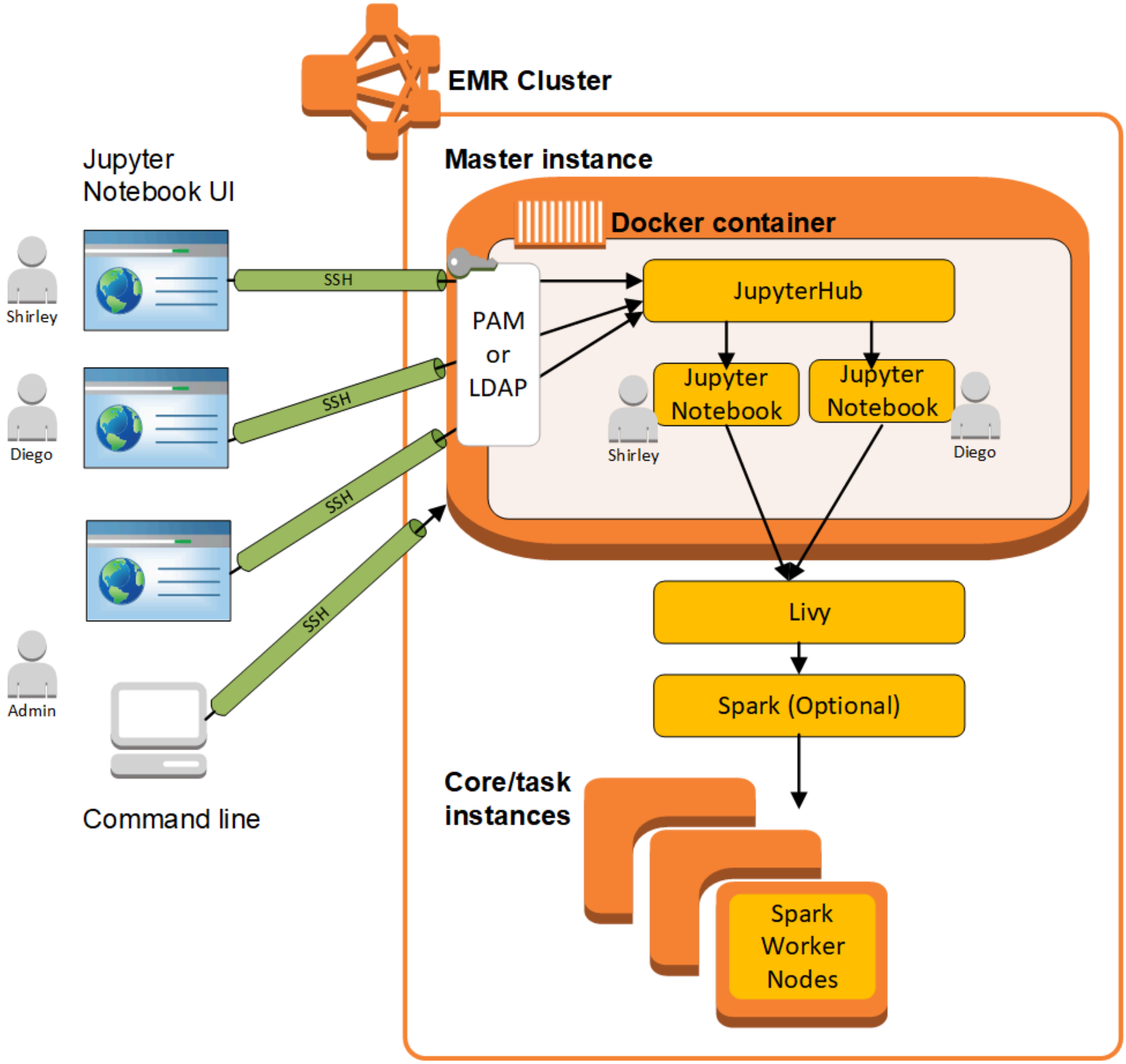
For more information, see EMR notebooks in the *Amazon EMR Management Guide*.

JupyterHub

Jupyter Notebook is an open-source web application that you can use to create and share documents that contain live code, equations, visualizations, and narrative text. JupyterHub allows you to host multiple instances of a single-user Jupyter notebook server. When you create a cluster with JupyterHub, Amazon EMR creates a Docker container on the cluster's master node. JupyterHub, all the components required for Jupyter, and Sparkmagic run within the container.

Sparkmagic is a library of kernels that allows Jupyter notebooks to interact with Apache Spark running on Amazon EMR through Apache Livy, which is a REST server for Spark. Spark and Apache Livy are installed automatically when you create a cluster with JupyterHub. The default Python 3 kernel for Jupyter is available along with the PySpark 3, PySpark, and Spark kernels that are available with Sparkmagic. You can use these kernels to run ad-hoc Spark code and interactive SQL queries using Python and Scala. You can install additional kernels within the Docker container manually. For more information, see Installing additional kernels and libraries.

The following diagram depicts the components of JupyterHub on Amazon EMR with corresponding authentication methods for notebook users and the administrator. For more information, see Adding Jupyter Notebook users and administrators.



The following table lists the version of JupyterHub included in the latest release of the Amazon EMR 7.x series, along with the components that Amazon EMR installs with JupyterHub.

For the version of components installed with JupyterHub in this release, see [Release 7.2.0 Component Versions](#).

JupyterHub version information for emr-7.2.0

Amazon EMR Release Label	JupyterHub Version	Components Installed With JupyterHub
emr-7.2.0	JupyterHub 1.5.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

The following table lists the version of JupyterHub included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with JupyterHub.

For the version of components installed with JupyterHub in this release, see Release 6.15.0 Component Versions.

JupyterHub version information for emr-6.15.0

Amazon EMR Release Label	JupyterHub Version	Components Installed With JupyterHub
emr-6.15.0	JupyterHub 1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resour

Amazon EMR Release Label	JupyterHub Version	Components Installed With JupyterHub
		cemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

The following table lists the version of JupyterHub included in the latest release of the Amazon EMR 5.x series, along with the components that Amazon EMR installs with JupyterHub.

For the version of components installed with JupyterHub in this release, see Release 5.36.2 Component Versions.

JupyterHub version information for emr-5.36.2

Amazon EMR Release Label	JupyterHub Version	Components Installed With JupyterHub
emr-5.36.2	JupyterHub 1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

The Python 3 kernel included with JupyterHub on Amazon EMR is 3.6.4.

The libraries installed within the `jupyterhub` container may vary between Amazon EMR release versions and Amazon EC2 AMI versions.

To list installed libraries using conda

- Run the following command on the master node command line:

```
sudo docker exec jupyterhub bash -c "conda list"
```

To list installed libraries using pip

- Run the following command on the master node command line:

```
sudo docker exec jupyterhub bash -c "pip freeze"
```

Topics

Create a cluster with JupyterHub

You can create an Amazon EMR cluster with JupyterHub using the AWS Management Console, AWS Command Line Interface, or the Amazon EMR API. Ensure that the cluster is not created with the option to terminate automatically after completing steps (`--auto-terminate` option in the AWS CLI). Also, make sure that administrators and notebook users can access the key pair that you use when you create the cluster. For more information, see Use a key pair for SSH credentials in the *Amazon EMR Management Guide*.

Create a cluster with JupyterHub using the console

Use the following procedure to create a cluster with JupyterHub installed using **Advanced Options** in the Amazon EMR console.

To create an Amazon EMR cluster with JupyterHub installed using the Amazon EMR console

1. Navigate to the new Amazon EMR console and select **Switch to the old console** from the side navigation. For more information on what to expect when you switch to the old console, see Using the old console.
2. Choose **Create cluster**, **Go to advanced options**.
3. Under **Software Configuration**:
   - For **Release**, select emr-5.36.2, and choose JupyterHub.
   - If you use Spark, to use the AWS Glue Data Catalog as the metastore for Spark SQL, select **Use for Spark table metadata**. For more information, see Use the AWS Glue Data Catalog as the metastore for Spark SQL.
   - For **Edit software settings** choose **Enter configuration** and specify values, or choose **Load JSON from S3** and specify a JSON configuration file. For more information, see Configuring JupyterHub.
4. Under **Add steps (optional)** configure steps to run when the cluster is created, make sure that **Auto-terminate cluster after the last step is completed** is not selected, and choose **Next**.
5. Choose **Hardware Configuration** options, **Next**. For more information, see Configure cluster hardware and networking in the *Amazon EMR Management Guide*.
6. Choose options for **General Cluster Settings**, **Next**.
7. Choose **Security Options**, specifying a key pair, and choose **Create Cluster**.

Create a cluster with JupyterHub using the AWS CLI

To launch a cluster with JupyterHub, use the `aws emr create-cluster` command and, for the `--applications` option, specify `Name=JupyterHub`. The following example launches a JupyterHub cluster on Amazon EMR with two EC2 instances (one master and one core instance). Also, debugging is enabled, with logs stored in the Amazon S3 location as specified by `--log-uri`. The specified key pair provides access to Amazon EC2 instances in the cluster.

Note

Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --name="MyJupyterHubCluster" --release-label emr-5.36.2 \
--applications Name=JupyterHub --log-uri s3://MyBucket/MyJupyterClusterLogs \
--use-default-roles --instance-type m5.xlarge --instance-count 2 --ec2-attributes
 KeyName=MyKeyPair
```

Considerations when using JupyterHub on Amazon EMR

Consider the following when using JupyterHub on Amazon EMR.

- **Warning**

  User notebooks and files are saved to the file system on the master node. This is ephemeral storage that does not persist through cluster termination. When a cluster terminates, this data is lost if not backed up. We recommend that you schedule regular backups using `cron` jobs or another means suitable for your application.
  In addition, configuration changes made within the container may not persist if the container restarts. We recommend that you script or otherwise automate container configuration so that you can reproduce customizations more readily.
- Kerberos authentication that has been set up using an Amazon EMR security configuration is not supported.
- OAuthenticator is not supported.

Configuring JupyterHub

You can customize the configuration of JupyterHub on Amazon EMR and individual user notebooks by connecting to the cluster master node and editing configuration files. After you change values, restart the `jupyterhub` container.

Modify properties in the following files to configure JupyterHub and individual Jupyter notebooks:

- jupyterhub_config.py—By default, this file is saved in the /etc/jupyter/conf/ directory on the master node. For more information, see Configuration basics in the JupyterHub documentation.
- jupyter_notebook_config.py—This file is saved in the /etc/jupyter/ directory by default and copied to the jupyterhub container as the default. For more information, see Config file and command line options in the Jupyter Notebook documentation.

You can also use the jupyter-sparkmagic-conf configuration classification to customize Sparkmagic, which updates values in the config.json file for Sparkmagic. For more information about available settings, see the example_config.json on GitHub. For more information about using configuration classifications with applications in Amazon EMR, see Configure applications.

The following example launches a cluster using the AWS CLI, referencing the file MyJupyterConfig.json for Sparkmagic configuration classification settings.

Note

Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --use-default-roles --release-label emr-5.14.0 \
--applications Name=Jupyter --instance-type m4.xlarge --instance-count 3 \
--ec2-attributes KeyName=MyKey,SubnetId=subnet-1234a5b6 --configurations file://
MyJupyterConfig.json
```

Sample contents of MyJupyterConfig.json are as follows:

```
[
    {
    "Classification":"jupyter-sparkmagic-conf",
    "Properties": {
      "kernel_python_credentials" : "{\"username\":\"diego\",\"base64_password\":
\"mypass\",\"url\":\"http:\/\/localhost:8998\",\"auth\":\"None\"}"
      }
    }
]
```

Note

With Amazon EMR version 5.21.0 and later, you can override cluster configurations and specify additional configuration classifications for each instance group in a running cluster. You do this by using the Amazon EMR console, the AWS Command Line Interface (AWS CLI), or the AWS SDK. For more information, see [Supplying a Configuration for an Instance Group in a Running Cluster](#).

Configuring persistence for notebooks in Amazon S3

You can configure a JupyterHub cluster in Amazon EMR so that notebooks saved by a user persist in Amazon S3, outside of ephemeral storage on cluster EC2 instances.

You specify Amazon S3 persistence using the `jupyter-s3-conf` configuration classification when you create a cluster. For more information, see Configure applications.

In addition to enabling Amazon S3 persistence using the `s3.persistence.enabled` property, you specify a bucket in Amazon S3 where notebooks are saved using the `s3.persistence.bucket` property. Notebooks for each user are saved to a `jupyter/`*`jupyterhub-user-name`* folder in the specified bucket. The bucket must already exist in Amazon S3, and the role for the EC2 instance profile that you specify when you create the cluster must have permissions to the bucket (by default, the role is `EMR_EC2_DefaultRole`). For more information, see Configure IAM roles for Amazon EMR permissions to AWS services.

When you launch a new cluster using the same configuration classification properties, users can open notebooks with the content from the saved location.

Note that when you import files as modules in a notebook when you have Amazon S3 enabled, this will result in the files uploading to Amazon S3. When you import files without enabling Amazon S3 persistence, they upload to your JupyterHub container.

The following example enables Amazon S3 persistence. Notebooks saved by users are saved in the `s3://MyJupyterBackups/jupyter/`*`jupyterhub-user-name`* folder for each user, where *`jupyterhub-user-name`* is a user name, such as `diego`.

```
[
    {
```

```
        "Classification": "jupyter-s3-conf",
        "Properties": {
            "s3.persistence.enabled": "true",
            "s3.persistence.bucket": "MyJupyterBackups"
        }
    }
]
```

Connecting to the master node and Notebook servers

JupyterHub administrators and notebook users must connect to the cluster master node using an SSH tunnel and then connecting to web interfaces served by JupyterHub on the master node. For more information about configuring an SSH tunnel and using the tunnel to proxy Web connections, see Connect to the cluster in the *Amazon EMR Management Guide*.

By default, JupyterHub on Amazon EMR is available through **port 9443** on the master node. The internal JupyterHub proxy also serves notebook instances through port 9443. JupyterHub and Jupyter web interfaces can be accessed using a URL with the following pattern:

https://*MasterNodeDNS*:9443

You can specify a different port using the `c.JupyterHub.port` property in the `jupyterhub_config.py` file. For more information, see Networking basics in the JupyterHub documentation.

By default, JupyterHub on Amazon EMR uses a self-signed certificate for SSL encryption using HTTPS. Users are prompted to trust the self-signed certificate when they connect. You can use a trusted certificate and keys of your own. Replace the default certificate file, `server.crt`, and key file `server.key` in the `/etc/jupyter/conf/` directory on the master node with certificate and key files of your own. Use the `c.JupyterHub.ssl_key` and `c.JupyterHub.ssl_cert` properties in the `jupyterhub_config.py` file to specify your SSL materials. For more information, see Security settings in the JupyterHub documentation. After you update `jupyterhub_config.py`, restart the container.

JupyterHub configuration and administration

JupyterHub and related components run inside a Docker container named `jupyterhub` that runs the Ubuntu operating system. There are several ways for you to administer components running inside the container.

> **Warning**
>
> Customizations that you perform within the container may not persist if the container restarts. We recommend that you script or otherwise automate container configuration so that you can reproduce customizations more readily.

Administration using the command line

When connected to the master node using SSH, you can issue commands by using the Docker command-line interface (CLI) and specifying the container by name (`jupyterhub`) or ID. For example, `sudo docker exec jupyterhub` *`command`* runs commands recognized by the operating system or an application running inside the container. You can use this method to add users to the operating system and to install additional applications and libraries within the Docker container. For example, the default container image includes Conda for package installation, so you might run the following command on the master node command line to install an application, Keras, within the container:

```
sudo docker exec jupyterhub conda install keras
```

Administration by submitting steps

Steps are a way to submit work to a cluster. You can submit steps when you launch a cluster, or you can submit steps to a running cluster. Commands that you run on the command line can be submitted as steps using `command-runner.jar`. For more information, see [Work with steps using the CLI and console] in the *Amazon EMR Management Guide* and [Run commands and scripts on an Amazon EMR cluster].

For example, you could use the following AWS CLI command on a local computer to install Keras in the same way that you did from the command line of the master node in the earlier example:

```
aws emr add-steps --cluster-id MyClusterID --steps Name="Command
 Runner",Jar="command-runner.jar",Args="/usr/bin/sudo","/usr/bin/
docker","exec","jupyterhub","conda","install","keras"
```

Also, you can script a sequence of steps, upload the script to Amazon S3, and then use `script-runner.jar` to run the script when you create the cluster or add the script as a step. For more information, see Run commands and scripts on an Amazon EMR cluster. For an example, see the section called "Example: Bash script to add multiple users".

Administration using REST APIs

Jupyter, JupyterHub, and the HTTP proxy for JupyterHub provide REST APIs that you can use to send requests. To send requests to JupyterHub, you must pass an API token with the request. You can use the `curl` command from the master node command line to execute REST commands. For more information, see the following resources:

- Using JupyterHub's REST API in the documentation for JupyterHub, which includes instructions for generating API tokens
- Jupyter Notebook server API on GitHub
- configurable-http-proxy on GitHub

The following example demonstrates using the REST API for JupyterHub to get a list of users. The command passes a previously generated admin token and uses the default port, 9443, for JupyterHub, piping the output to jq for easier viewing:

```
curl -XGET -s -k https://$HOST:9443/hub/api/users \
-H "Authorization: token $admin_token" | jq .
```

Adding Jupyter Notebook users and administrators

You can use one of two methods for users to authenticate to JupyterHub so that they can create notebooks and, optionally, administer JupyterHub. The easiest method is to use JupyterHub's pluggable authentication module (PAM). In addition, JupyterHub on Amazon EMR supports the LDAP authenticator plugin for JupyterHub for obtaining user identities from an LDAP server, such as a Microsoft Active Directory server. Instructions and examples for adding users with each authentication method are provided in this section.

JupyterHub on Amazon EMR has a default user with administrator permissions. The user name is `jovyan` and the password is `jupyter`. We strongly recommend that you replace the user with another user who has administrative permissions. You can do this using a step when you create the cluster, or by connecting to the master node when the cluster is running.

Topics

- Using LDAP authentication
- User impersonation

Using PAM authentication

Creating PAM users in JupyterHub on Amazon EMR is a two-step process. The first step is to add users to the operating system running in the `jupyterhub` container on the master node, and to add a corresponding user home directory for each user. The second step is to add these operating system users as JupyterHub users—a process known as *whitelisting* in JupyterHub. After a JupyterHub user is added, they can connect to the JupyterHub URL and provide their operating system credentials for access.

When a user logs in, JupyterHub opens the notebook server instance for that user, which is saved in the user's home directory on the master node, which is `/var/lib/jupyter/home/username`. If a notebook server instance doesn't exist, JupyterHub spawns a notebook instance in the user's home directory. The following sections demonstrate how to add users individually to the operating system and to JupyterHub, followed by a rudimentary bash script that adds multiple users.

Adding an operating system user to the container

The following example first uses the useradd command within the container to add a single user, diego, and create a home directory for that user. The second command uses chpasswd to establish a password of diego for this user. Commands are run on the master node command line while connected using SSH. You could also run these commands using a step as described earlier in Administration by submitting steps.

```
sudo docker exec jupyterhub useradd -m -s /bin/bash -N diego
sudo docker exec jupyterhub bash -c "echo diego:diego | chpasswd"
```

Adding a JupyterHub user

You can use the **Admin** panel in JupyterHub or the REST API to add users and administrators, or just users.

To add users and administrators using the admin panel in JupyterHub

1. Connect to the master node using SSH and log in to https://*MasterNodeDNS*:9443 with an identity that has administrator permissions.

2. Choose **Control Panel**, **Admin**.
3. Choose **User**, **Add Users**, or choose **Admin**, **Add Admins**.

To add a user using the REST API

1. Connect to the master node using SSH and use the following command on the master node, or run the command as a step.
2. Acquire an administrative token to make API requests, and replace *AdminToken* in the following step with that token.
3. Use the following command, replacing *UserName* with an operating system user that has been created within the container.

```
curl -XPOST -H "Authorization: token AdminToken" "https://$(hostname):9443/hub/api/
users/UserName
```

Note

You are automatically added as a JupyterHub non-admin user when you log in to the JupyterHub web interface for the first time.

Example: Bash script to add multiple users

The following sample bash script ties together the previous steps in this section to create multiple JupyterHub users. The script can be run directly on the master node, or it can be uploaded to Amazon S3 and then run as a step.

The script first establishes an array of user names, and uses the `jupyterhub token` command to create an API token for the default administrator, jovyan. It then creates an operating system user in the `jupyterhub` container for each user, assigning an initial password to each that is equal to their user name. Finally, it calls the REST API operation to create each user in JupyterHub. It passes the token generated earlier in the script and pipes the REST response to `jq` for easier viewing.

```
# Bulk add users to container and JupyterHub with temp password of username
set -x
USERS=(shirley diego ana richard li john mary anaya)
TOKEN=$(sudo docker exec jupyterhub /opt/conda/bin/jupyterhub token jovyan | tail -1)
```

```
for i in "${USERS[@]}";
do
   sudo docker exec jupyterhub useradd -m -s /bin/bash -N $i
   sudo docker exec jupyterhub bash -c "echo $i:$i | chpasswd"
   curl -XPOST --silent -k https://$(hostname):9443/hub/api/users/$i \
 -H "Authorization: token $TOKEN" | jq
done
```

Save the script to a location in Amazon S3 such as `s3://mybucket/createjupyterusers.sh`. Then you can use `script-runner.jar` to run it as a step.

Example: Running the script when creating a cluster (AWS CLI)

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --name="MyJupyterHubCluster" --release-label emr-5.36.2 \
--applications Name=JupyterHub --log-uri s3://MyBucket/MyJupyterClusterLogs \
--use-default-roles --instance-type m5.xlarge --instance-count 2 --ec2-attributes
 KeyName=MyKeyPair \
--steps Type=CUSTOM_JAR,Name=CustomJAR,ActionOnFailure=CONTINUE,\
Jar=s3://region.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://
mybucket/createjupyterusers.sh"]
```

Running the script on an existing cluster (AWS CLI)

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr add-steps --cluster-id j-XXXXXXXX --steps Type=CUSTOM_JAR,\
Name=CustomJAR,ActionOnFailure=CONTINUE,\
Jar=s3://region.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://
mybucket/createjupyterusers.sh"]
```

Using LDAP authentication

Lightweight Directory Access Protocol (LDAP) is an application protocol for querying and modifying objects that correspond to resources such as users and computers stored in an LDAP-compatible directory service provider such as Active Directory or an OpenLDAP server. You can use the LDAP authenticator plugin for JupyterHub with JupyterHub on Amazon EMR to use LDAP for user authentication. The plugin handles login sessions for LDAP users and provides user information to Jupyter. This lets users connect to JupyterHub and notebooks by using the credentials for their identities stored in an LDAP-compatible server.

The steps in this section walk you through the following steps to set up and enable LDAP using the LDAP Authenticator Plugin for JupyterHub. You perform the steps while connected to the master node command line. For more information, see Connecting to the master node and Notebook servers.

1. Create an LDAP configuration file with information about the LDAP server, such as the host IP address, port, binding names, and so on.
2. Modify `/etc/jupyter/conf/jupyterhub_config.py` to enable the LDAP Authenticator Plugin for JupyterHub.
3. Create and run a script that configures LDAP within the `jupyterhub` container.
4. Query LDAP for users, and then create home directories within the container for each user. JupyterHub requires home directories to host notebooks.
5. Run a script that restarts JupyterHub

> **Important**
>
> Before you set up LDAP, test your network infrastructure to ensure that the LDAP server and the cluster master node can communicate as required. TLS typically uses port 389 over a plain TCP connection. If your LDAP connection uses SSL, the well-known TCP port for SSL is 636.

Create the LDAP configuration file

The example below uses the following place-holder configuration values. Replace these with parameters that match your implementation.

- The LDAP server is running version 3 and available on port 389. This is the standard non-SSL port for LDAP.
- The base distinguished name (DN) is `dc=example, dc=org`.

Use a text editor to create the file ldap.conf, with contents similar to the following. Use values appropriate for your LDAP implementation. Replace *host* with the IP address or resolvable host name of your LDAP server.

```
base dc=example,dc=org
uri ldap://host
ldap_version 3
binddn cn=admin,dc=example,dc=org
bindpw admin
```

Enable LDAP Authenticator Plugin for JupyterHub

Use a text editor to modify the `/etc/jupyter/conf/jupyterhub_config.py` file and add ldapauthenticator properties similar to the following. Replace *host* with the IP address or resolvable host name of the LDAP server. The example assumes that the user objects are within an organizational unit (ou) named *people*, and uses the distinguished name components that you established earlier using `ldap.conf`.

```
c.JupyterHub.authenticator_class = 'ldapauthenticator.LDAPAuthenticator'
c.LDAPAuthenticator.use_ssl = False
c.LDAPAuthenticator.server_address = 'host'
c.LDAPAuthenticator.bind_dn_template = 'cn={username},ou=people,dc=example,dc=org'
```

Configure LDAP within the container

Use a text editor to create a bash script with the following contents:

```
#!/bin/bash

# Uncomment the following lines to install LDAP client libraries only if
# using Amazon EMR release version 5.14.0. Later versions install libraries by default.
# sudo docker exec jupyterhub bash -c "sudo apt-get update"
# sudo docker exec jupyterhub bash -c "sudo apt-get -y install libnss-ldap libpam-ldap
 ldap-utils nscd"

# Copy ldap.conf
```

```
sudo docker cp ldap.conf jupyterhub:/etc/ldap/
sudo docker exec jupyterhub bash -c "cat /etc/ldap/ldap.conf"

# configure nss switch
sudo docker exec jupyterhub bash -c "sed -i 's/\(^passwd.*\)/\1 ldap/g' /etc/
nsswitch.conf"
sudo docker exec jupyterhub bash -c "sed -i 's/\(^group.*\)/\1 ldap/g' /etc/
nsswitch.conf"
sudo docker exec jupyterhub bash -c "sed -i 's/\(^shadow.*\)/\1 ldap/g' /etc/
nsswitch.conf"
sudo docker exec jupyterhub bash -c "cat /etc/nsswitch.conf"

# configure PAM to create home directories
sudo docker exec jupyterhub bash -c "echo 'session required        pam_mkhomedir.so
 skel=/etc/skel umask=077' >> /etc/pam.d/common-session"
sudo docker exec jupyterhub bash -c "cat /etc/pam.d/common-session"

# restart nscd service
sudo docker exec jupyterhub bash -c "sudo service nscd restart"

# Test
sudo docker exec jupyterhub bash -c "getent passwd"

# Install ldap plugin
sudo docker exec jupyterhub bash -c "pip install jupyterhub-ldapauthenticator"
```

Save the script to the master node, and then run it from the master node command line. For example, with the script saved as `configure_ldap_client.sh`, make the file executable:

```
chmod +x configure_ldap_client.sh
```

And run the script:

```
./configure_ldap_client.sh
```

Add attributes to Active Directory

To find each user and create the appropriate entry in the database, the JupyterHub docker container requires the following UNIX properties for the corresponding user object in Active Directory. For more information, see the section *How do I continue to edit the GID/UID RFC 2307 attributes now that the Unix Attributes Plug-in is no longer available for the Active Directory Users and Computers MMC snap-in?* in the article Clarification regarding the status of identity management for Unix (IDMU) and NIS server role in Windows Server 2016 technical preview and beyond.

- `homeDirectory`

  This is the location to the user's home directory, which is usually `/home/username`.
- `gidNumber`

  This is a value greater than 60000 that is not already used by a another user. Check the `etc/passwd` file for gids in use.
- `uidNumber`

  This is a value greater than 60000 that is not already used by a another group. Check the `etc/group` file for uids in use.
- `uid`

  This is the same as the *username*.

Create user home directories

JupyterHub needs home directories within the container to authenticate LDAP users and store instance data. The following example demonstrates two users, *shirley* and *diego*, in the LDAP directory.

The first step is to query the LDAP server for each user's user id and group id information using ldapsearch as shown in the following example, replacing *host* with the IP address or resolvable host name of your LDAP server:

```
ldapsearch -x -H ldap://host \
 -D "cn=admin,dc=example,dc=org" \
 -w admin \
 -b "ou=people,dc=example,dc=org" \
 -s sub \
 "(objectclass=*)" uidNumber gidNumber
```

The `ldapsearch` command returns an LDIF-formatted response that looks similar to the following for users *shirley* and *diego*.

```
# extended LDIF
```

```
# LDAPv3
# base <ou=people,dc=example,dc=org> with scope subtree
# filter: (objectclass=*)
# requesting: uidNumber gidNumber sn

# people, example.org
dn: ou=people,dc=example,dc=org

# diego, people, example.org
dn: cn=diego,ou=people,dc=example,dc=org
sn: B
uidNumber: 1001
gidNumber: 100

# shirley, people, example.org
dn: cn=shirley,ou=people,dc=example,dc=org
sn: A
uidNumber: 1002
gidNumber: 100

# search result
search: 2
result: 0 Success

# numResponses: 4
# numEntries: 3
```

Using information from the response, run commands within the container to create a home directory for each user common name (`cn`). Use the `uidNumber` and `gidNumber` to fix ownership for the home directory for that user. The following example commands do this for the user *`shirley`*.

```
sudo docker container exec jupyterhub bash -c "mkdir /home/shirley"
sudo docker container exec jupyterhub bash -c "chown -R $uidNumber /home/shirley"
sudo docker container exec jupyterhub bash -c "sudo chgrp -R $gidNumber /home/shirley"
```

Note

LDAP authenticator for JupyterHub does not support local user creation. For more information, see [LDAP authenticator configuration note on local user creation](#).
To create a local user manually, use the following command.

```
sudo docker exec jupyterhub bash -c "echo 'shirley:x:$uidNumber:$gidNumber::/
home/shirley:/bin/bash' >> /etc/passwd"
```

Restart the JupyterHub container

Run the following commands to restart the `jupyterhub` container:

```
sudo docker stop jupyterhub
sudo docker start jupyterhub
```

User impersonation

A Spark job running inside a Jupyter notebook traverses multiple applications during its execution on Amazon EMR. For example, PySpark3 code that a user runs inside Jupyter is received by Sparkmagic, which uses an HTTP POST request to submit it to Livy, which then creates a Spark job to execute on the cluster using YARN.

By default, YARN jobs submitted this way run as user `livy`, regardless of the user who initiated the job. By setting up *user impersonation* you can have the user ID of the notebook user also be the user associated with the YARN job. Rather than having jobs initiated by both `shirley` and `diego` associated with the user `livy`, jobs that each user initiates are associated with `shirley` and `diego` respectively. This helps you to audit Jupyter usage and manage applications within your organization.

This configuration is only supported when calls from Sparkmagic to Livy are unauthenticated. Applications that provide an authentication or proxying layer between Hadoop applications and Livy (such as Apache Knox Gateway) are not supported. The steps to configure user impersonation in this section assume that JupyterHub and Livy are running on the same master node. If your application has separate clusters, Step 3: Create HDFS home directories for users needs to be modified so that HDFS directories are created on the Livy master node.

Steps to configure user impersonation

- Step 1: Configure Livy
- Step 2: Add users
- Step 3: Create HDFS home directories for users

Step 1: Configure Livy

You use the `livy-conf` and `core-site` configuration classifications when you create a cluster to enable Livy user impersonation as shown in the following example. Save the configuration classification as a JSON and then reference it when you create the cluster, or specify the configuration classification inline. For more information, see Configure applications.

```
[
  {
    "Classification": "livy-conf",
    "Properties": {
      "livy.impersonation.enabled": "true"
    }
  },
  {
    "Classification": "core-site",
    "Properties": {
      "hadoop.proxyuser.livy.groups": "*",
      "hadoop.proxyuser.livy.hosts": "*"
    }
  }
]
```

Step 2: Add users

Add JupyterHub users using PAM or LDAP. For more information, see Using PAM authentication and Using LDAP authentication.

Step 3: Create HDFS home directories for users

You connected to the master node to create users. While still connected to the master node, copy the contents below and save it to a script file. The script creates HDFS home directories for each JupyterHub user on the master node. The script assumes you are using the default administrator user ID, *jovyan*.

```
#!/bin/bash

CURL="curl --silent -k"
HOST=$(curl -s http://169.254.169.254/latest/meta-data/local-hostname)

admin_token() {
```

```
    local user=jovyan
    local pwd=jupyter
    local token=$($CURL https://$HOST:9443/hub/api/authorizations/token \
        -d "{\"username\":\"$user\", \"password\":\"$pwd\"}" | jq ".token")
    if [[ $token != null ]]; then
        token=$(echo $token | sed 's/"//g')
    else
        echo "Unable to get Jupyter API Token."
        exit 1
    fi
    echo $token
}

# Get Jupyter Admin token
token=$(admin_token)

# Get list of Jupyter users
users=$(curl -XGET -s -k https://$HOST:9443/hub/api/users \
 -H "Authorization: token $token" | jq '.[].name' | sed 's/"//g')

# Create HDFS home dir
for user in ${users[@]};
do
 echo "Create hdfs home dir for $user"
 hadoop fs -mkdir /user/$user
 hadoop fs -chmod 777 /user/$user
done
```

Installing additional kernels and libraries

When you create a cluster with JupyterHub on Amazon EMR, the default Python 3 kernel for Jupyter along with the PySpark and Spark kernels for Sparkmagic are installed on the Docker container. You can install additional kernels. You can also install additional libraries and packages and then import them for the appropriate shell.

Installing a kernel

Kernels are installed within the Docker container. The easiest way to accomplish this is to create a bash script with installation commands, save it to the master node, and then use the `sudo docker exec jupyterhub` *`script_name`* command to run the script within the `jupyterhub` container. The following example script installs the kernel, and then installs a few libraries for that kernel on the master node so that later you can import the libraries using the kernel in Jupyter.

```
#!/bin/bash

# Install Python 2 kernel
conda create -n py27 python=2.7 anaconda
source /opt/conda/envs/py27/bin/activate
apt-get update
apt-get install -y gcc
/opt/conda/envs/py27/bin/python -m pip install --upgrade ipykernel
/opt/conda/envs/py27/bin/python -m ipykernel install

# Install libraries for Python 2
/opt/conda/envs/py27/bin/pip install paramiko nltk scipy numpy scikit-learn pandas
```

To install the kernel and libraries within the container, open a terminal connection to the master node, save the script to `/etc/jupyter/install_kernels.sh`, and run the following command on the master node command line:

```
sudo docker exec jupyterhub bash /etc/jupyter/install_kernels.sh
```

Using libraries and installing additional libraries

A core set of machine learning and data science libraries for Python 3 are pre-installed with JupyterHub on Amazon EMR. You can use `sudo docker exec jupyterhub bash -c "conda list"` and `sudo docker exec jupyterhub bash -c "pip freeze"`.

If a Spark job needs libraries on worker nodes, we recommend that you use a bootstrap action to run a script to install the libraries when you create the cluster. Bootstrap actions run on all cluster nodes during the cluster creation process, which simplifies installation. If you install libraries on core/worker nodes after a cluster is running, the operation is more complex. We provide an example Python program in this section that shows how to install these libraries.

The bootstrap action and Python program examples shown in this section use a bash script saved to Amazon S3 to install the libraries on all nodes.

The script referenced in the following example uses `pip` to install paramiko, nltk, scipy, scikit-learn, and pandas for the Python 3 kernel:

```
#!/bin/bash
```

```
sudo python3 -m pip install boto3 paramiko nltk scipy scikit-learn pandas
```

After you create the script, upload it to a location in Amazon S3, for example, `s3://mybucket/install-my-jupyter-libraries.sh`. For more information, see Uploading objects in the *Amazon Simple Storage Service User Guide* so that you can use it in your bootstrap action or in your Python program.

To specify a bootstrap action that installs libraries on all nodes when you create a cluster using the AWS CLI

1. Create a script similar to the earlier example and save it to a location in Amazon S3. We use the example `s3://mybucket/install-my-jupyter-libraries.sh`.
2. Create the cluster with JupyterHub and use the `Path` argument of the `--bootstrap-actions` option to specify the script location as shown in the following example:

   Note

   Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --name="MyJupyterHubCluster" --release-label emr-5.36.2 \
--applications Name=JupyterHub --log-uri s3://MyBucket/MyJupyterClusterLogs \
--use-default-roles --instance-type m5.xlarge --instance-count 2 --ec2-attributes
 KeyName=MyKeyPair \
--bootstrap-actions Path=s3://mybucket/install-my-jupyter-
libraries.sh,Name=InstallJupyterLibs
```

To specify a bootstrap action that installs libraries on all nodes when you create a cluster using the console

1. Navigate to the new Amazon EMR console and select **Switch to the old console** from the side navigation. For more information on what to expect when you switch to the old console, see Using the old console.
2. Choose **Create cluster**, **Go to advanced options**.
3. Specify settings for **Software and Steps** and **Hardware** as appropriate for your application.
4. On the **General Cluster Settings** screen, expand **Bootstrap Actions**.

5. For **Add bootstrap action**, select **Custom action**, **Configure and add**.
6. For **Name**, enter a friendly name. For **Script location**, enter the location in Amazon S3 of your script (the example we use is *s3://mybucket/install-my-jupyter-libraries.sh*). Leave **Optional arguments** blank, and choose **Add**.
7. Specify other settings for your cluster, and choose **Next**.
8. Specify security settings, and choose **Create cluster**.

Example Installing libraries on core nodes of a running cluster

After you install libraries on the master node from within Jupyter, you can install libraries on running core nodes in various ways. The following example shows a Python program written to run on a local machine. When you run the Python program locally, it uses the `AWS-RunShellScript` of AWS Systems Manager to run the example script, shown earlier in this section, which installs libraries on the cluster's core nodes.

```
import argparse
import time
import boto3


def install_libraries_on_core_nodes(cluster_id, script_path, emr_client, ssm_client):
    """
    Copies and runs a shell script on the core nodes in the cluster.

    :param cluster_id: The ID of the cluster.
    :param script_path: The path to the script, typically an Amazon S3 object URL.
    :param emr_client: The Boto3 Amazon EMR client.
    :param ssm_client: The Boto3 AWS Systems Manager client.
    """
    core_nodes = emr_client.list_instances(
        ClusterId=cluster_id, InstanceGroupTypes=["CORE"]
    )["Instances"]
    core_instance_ids = [node["Ec2InstanceId"] for node in core_nodes]
    print(f"Found core instances: {core_instance_ids}.")

    commands = [
        # Copy the shell script from Amazon S3 to each node instance.
        f"aws s3 cp {script_path} /home/hadoop",
        # Run the shell script to install libraries on each node instance.
        "bash /home/hadoop/install_libraries.sh",
```

```
    ]
    for command in commands:
        print(f"Sending '{command}' to core instances...")
        command_id = ssm_client.send_command(
            InstanceIds=core_instance_ids,
            DocumentName="AWS-RunShellScript",
            Parameters={"commands": [command]},
            TimeoutSeconds=3600,
        )["Command"]["CommandId"]
        while True:
            # Verify the previous step succeeded before running the next step.
            cmd_result = ssm_client.list_commands(CommandId=command_id)["Commands"][0]
            if cmd_result["StatusDetails"] == "Success":
                print(f"Command succeeded.")
                break
            elif cmd_result["StatusDetails"] in ["Pending", "InProgress"]:
                print(f"Command status is {cmd_result['StatusDetails']}, waiting...")
                time.sleep(10)
            else:
                print(f"Command status is {cmd_result['StatusDetails']}, quitting.")
                raise RuntimeError(
                    f"Command {command} failed to run. "
                    f"Details: {cmd_result['StatusDetails']}"
                )


def main():
    parser = argparse.ArgumentParser()
    parser.add_argument("cluster_id", help="The ID of the cluster.")
    parser.add_argument("script_path", help="The path to the script in Amazon S3.")
    args = parser.parse_args()

    emr_client = boto3.client("emr")
    ssm_client = boto3.client("ssm")

    install_libraries_on_core_nodes(
        args.cluster_id, args.script_path, emr_client, ssm_client
    )


if __name__ == "__main__":
    main()
```

JupyterHub release history

The following table lists the version of JupyterHub included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in Amazon EMR 7.x release versions, Amazon EMR 6.x release versions, or Amazon EMR 5.x release versions.

JupyterHub version information

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-7.2.0	1.5.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.36.2	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn,

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
		spark-yarn-slave, livy-server, jupyterhub
emr-7.1.0	1.5.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-7.0.0	1.5.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-6.15.0	1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.14.0	1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-6.13.0	1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.12.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-6.11.1	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.11.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-6.10.1	1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.10.0	1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-6.9.1	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.9.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-6.8.1	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.8.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-6.7.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.36.1	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-5.36.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.6.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-5.35.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.5.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-6.4.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.3.1	1.2.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-6.3.0	1.2.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.2.1	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-6.2.0	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.1.1	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-6.1.0	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.0.1	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-6.0.0	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.34.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-5.33.1	1.2.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.33.0	1.2.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-5.32.1	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.32.0	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-5.31.1	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.31.0	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-5.30.2	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.30.1	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-5.30.0	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.29.0	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-5.28.1	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.28.0	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-5.27.1	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.27.0	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-5.26.0	0.9.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.25.0	0.9.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-5.24.1	0.9.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.24.0	0.9.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-5.23.1	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.23.0	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-5.22.0	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.21.2	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-5.21.1	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.21.0	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-5.20.1	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.20.0	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-5.19.1	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.19.0	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-5.18.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.18.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-5.17.2	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.17.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-5.17.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.16.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-5.16.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.15.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-5.15.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.14.2	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon EMR Release label	JupyterHub Version	Components installed with JupyterHub
emr-5.14.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.14.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Apache Livy

Livy enables interaction over a REST interface with an EMR cluster running Spark. You can use the REST interface or an RPC client library to submit Spark jobs or snippets of Spark code, retrieve results synchronously or asynchronously, and manage Spark Context. For more information, see the Apache Livy website. Livy is included in Amazon EMR release version 5.9.0 and later.

To access the Livy web interface, set up an SSH tunnel to the master node and a proxy connection. For more information, see View web interfaces hosted on EMR clusters.

The following table lists the version of Livy included in the latest release of the Amazon EMR 7.x series, along with the components that Amazon EMR installs with Livy.

For the version of components installed with Livy in this release, see Release 7.2.0 Component Versions.

Livy version information for emr-7.2.0

Amazon EMR Release Label	Livy Version	Components Installed With Livy
emr-7.2.0	Livy 0.8.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

The following table lists the version of Livy included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with Livy.

For the version of components installed with Livy in this release, see Release 6.15.0 Component Versions.

Livy version information for emr-6.15.0

Amazon EMR Release Label	Livy Version	Components Installed With Livy
emr-6.15.0	Livy 0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

The following table lists the version of Livy included in the latest release of the Amazon EMR 5.x series, along with the components that Amazon EMR installs with Livy.

For the version of components installed with Livy in this release, see Release 5.36.2 Component Versions.

Livy version information for emr-5.36.2

Amazon EMR Release Label	Livy Version	Components Installed With Livy
emr-5.36.2	Livy 0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-

Amazon EMR Release Label	Livy Version	Components Installed With Livy
		namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Topics

Enabling HTTPS with Apache Livy

1. Provision an Amazon EMR cluster with transit encryption enabled. To learn more about encryption, see Encrypt data at rest and in transit.
2. Create a file called `livy_ssl.sh` with the following contents.

```
#!/bin/bash

KEYSTORE_FILE=`awk '/ssl.server.keystore.location/{getline; print}' /etc/hadoop/
conf/ssl-server.xml | sed -e 's/<[^>]*>//g' | tr -d ' \t\n\r\f'`
KEYSTORE_PASS=`awk '/ssl.server.keystore.password/{getline; print}' /etc/hadoop/
conf/ssl-server.xml | sed -e 's/<[^>]*>//g' | tr -d ' \t\n\r\f'`
KEY_PASS=`awk '/ssl.server.keystore.keypassword/{getline; print}' /etc/hadoop/conf/
ssl-server.xml | sed -e 's/<[^>]*>//g' | tr -d ' \t\n\r\f'`

echo "livy.keystore $KEYSTORE_FILE
livy.keystore.password $KEYSTORE_PASS
livy.key-password $KEY_PASS" | sudo tee -a /etc/livy/conf/livy.conf >/dev/null
```

```
sudo systemctl restart livy-server.service
```

3. Run the following script as an Amazon EMR step. This script modifies `/etc/livy/conf/livy.conf` to activate SSL.

```
--steps '[{"Args":["s3://DOC-EXAMPLE-BUCKET/
livy_ssl.sh"],"Type":"CUSTOM_JAR","ActionOnFailure":"CONTINUE","Jar":"s3://
us-east-1.elasticmapreduce/libs/script-runner/script-
runner.jar","Properties":"","Name":"Custom JAR"}]'
```

4. Restart the Apache Livy service so that the changes take effect. To restart Apache Livy, see Stopping and restarting processes.

5. Test that the clients can now communicate using HTTPS. To submit a job, for example, run the following code.

```
curl -k -X POST --data '{"file": "local:///usr/lib/spark/examples/jars/spark-
examples.jar",
"className": "org.apache.spark.examples.SparkPi"}' \
-H "Content-Type: application/json" \
https://EMR_Master_Node_Host:8998/batches
```

If you've enabled HTTPS successfully, Livy sends a response indicating that the command was accepted and that the batch job was submitted.

```
{"id":1,"name":null,"owner":null,"proxyUser":null,"state":"starting","appId":null,"appInfo":
{"driverLogUrl":null,"sparkUiUrl":null},"log":["stdout: ","\nstderr: ","\nYARN
 Diagnostics: "]}
```

Livy release history

The following table lists the version of Livy included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in Amazon EMR 7.x release versions, Amazon EMR 6.x release versions, or Amazon EMR 5.x release versions.

Livy version information

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-7.2.0	0.8.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.36.2	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-7.1.0	0.8.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-

Amazon EMR Release label	Livy Version	Components installed with Livy
		namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-7.0.0	0.7.1	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-6.15.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.14.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-6.13.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.12.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-6.11.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.11.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-6.10.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.10.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-6.9.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.9.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-6.8.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.8.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-6.7.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.36.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.36.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.6.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.35.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.5.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-6.4.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.3.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-6.3.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.2.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-6.2.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.1.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-6.1.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.0.1	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-6.0.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.34.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.33.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.33.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.32.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.32.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.31.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.31.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.30.2	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.30.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.30.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-notebook-env, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.29.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.28.1	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.28.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.27.1	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.27.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.26.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.25.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.24.1	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.24.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.23.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.23.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.22.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.21.2	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.21.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.21.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.20.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.20.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.19.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.19.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.18.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.18.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.17.2	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.17.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.17.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.16.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.16.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.15.1	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.15.0	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.14.2	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.14.1	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.14.0	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.13.1	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.13.0	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.12.3	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.12.2	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.12.1	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.12.0	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.11.4	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.11.3	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.11.2	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.11.1	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.11.0	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.10.1	0.4.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.10.0	0.4.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.9.1	0.4.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Amazon EMR Release label	Livy Version	Components installed with Livy
emr-5.9.0	0.4.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Apache MXNet

Apache MXNet is an acceleration library designed for building neural networks and other deep learning applications. MXNet automates common work flows and optimizes numerical computations. MXNet helps you design neural network architectures without having to focus on implementing low-level computations, such as linear algebra operations. MXNet is included with Amazon EMR release version 5.10.0 and later.

For more information, see the Apache MXNet web site.

The following table lists the version of MXNet included in the latest release of the Amazon EMR 7.x series, along with the components that Amazon EMR installs with MXNet.

For the version of components installed with MXNet in this release, see Release 7.2.0 Component Versions.

MXNet version information for emr-7.2.0

Amazon EMR Release Label	MXNet Version	Components Installed With MXNet
emr-7.2.0	MXNet 1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

The following table lists the version of MXNet included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with MXNet.

For the version of components installed with MXNet in this release, see Release 6.15.0 Component Versions.

MXNet version information for emr-6.15.0

Amazon EMR Release Label	MXNet Version	Components Installed With MXNet
emr-6.15.0	MXNet 1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

The following table lists the version of MXNet included in the latest release of the Amazon EMR 5.x series, along with the components that Amazon EMR installs with MXNet.

For the version of components installed with MXNet in this release, see Release 5.36.2 Component Versions.

MXNet version information for emr-5.36.2

Amazon EMR Release Label	MXNet Version	Components Installed With MXNet
emr-5.36.2	MXNet 1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

MXNet release history

The following table lists the version of MXNet included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in Amazon EMR 7.x release versions, Amazon EMR 6.x release versions, or Amazon EMR 5.x release versions.

MXNet version information

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-7.2.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.36.2	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-7.1.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-librar

Amazon EMR Release label	MXNet Version	Components installed with MXNet
		y, hadoop-hdfs-nameno de, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-7.0.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datano de, hadoop-hdfs-librar y, hadoop-hdfs-nameno de, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.15.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datano de, hadoop-hdfs-librar y, hadoop-hdfs-nameno de, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-6.14.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.13.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.12.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-6.11.1	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.11.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.10.1	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-6.10.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.9.1	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.9.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-6.8.1	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.8.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.7.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-5.36.1	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.36.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.6.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-5.35.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.5.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.4.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-6.3.1	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.3.0	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.2.1	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-6.2.0	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.1.1	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.1.0	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-6.0.1	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.0.0	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.34.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-5.33.1	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.33.0	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.32.1	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-5.32.0	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.31.1	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.31.0	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-5.30.2	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.30.1	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.30.0	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-5.29.0	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.28.1	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.28.0	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-5.27.1	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.27.0	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.26.0	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-5.25.0	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.24.1	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.24.0	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-5.23.1	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.23.0	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.22.0	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-5.21.2	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.21.1	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.21.0	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-5.20.1	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.20.0	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.19.1	1.3.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-5.19.0	1.3.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.18.1	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.18.0	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-5.17.2	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.17.1	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.17.0	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-5.16.1	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.16.0	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.15.1	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-5.15.0	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.14.2	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.14.1	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-5.14.0	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.13.1	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.13.0	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-5.12.3	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.12.2	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.12.1	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-5.12.0	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.11.4	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.11.3	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-5.11.2	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.11.1	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.11.0	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

Amazon EMR Release label	MXNet Version	Components installed with MXNet
emr-5.10.1	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.10.0	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

Apache Oozie

Use the Apache Oozie Workflow Scheduler to manage and coordinate Hadoop jobs. For more information, see http://oozie.apache.org/.

The Oozie native web interface is not supported on Amazon EMR. To use a front-end interface for Oozie, try the Hue Oozie application. For more information, see Hue. Oozie is included with Amazon EMR release version 5.0.0 and later. Oozie is included as a sandbox application in earlier releases. For more information, see Amazon EMR 4.x release versions.

If you use a custom Amazon Linux AMI based on an Amazon Linux AMI with a creation date of 2018-08-11, the Oozie server fails to start. If you use Oozie, create a custom AMI based on an Amazon Linux AMI ID with a different creation date. You can use the following AWS CLI command to return a list of Image IDs for all HVM Amazon Linux AMIs with a 2018.03 version, along with the release date, so that you can choose an appropriate Amazon Linux AMI as your base. Replace MyRegion with your Region identifier, such as us-west-2.

```
aws ec2 --region MyRegion describe-images --owner amazon --query 'Images[?
Name!=`null`]|[?starts_with(Name, `amzn-ami-hvm-2018.03`) == `true`].
[CreationDate,ImageId,Name]' --output text | sort -rk1
```

The following table lists the version of Oozie included in the latest release of the Amazon EMR 7.x series, along with the components that Amazon EMR installs with Oozie.

For the version of components installed with Oozie in this release, see Release 7.2.0 Component Versions.

Oozie version information for emr-7.2.0

Amazon EMR Release Label	Oozie Version	Components Installed With Oozie
emr-7.2.0	Oozie 5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-

Amazon EMR Release Label	Oozie Version	Components Installed With Oozie
		server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

The following table lists the version of Oozie included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with Oozie.

For the version of components installed with Oozie in this release, see Release 6.15.0 Component Versions.

Oozie version information for emr-6.15.0

Amazon EMR Release Label	Oozie Version	Components Installed With Oozie
emr-6.15.0	Oozie 5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

The following table lists the version of Oozie included in the latest release of the Amazon EMR 5.x series, along with the components that Amazon EMR installs with Oozie.

For the version of components installed with Oozie in this release, see Release 5.36.2 Component Versions.

Oozie version information for emr-5.36.2

Amazon EMR Release Label	Oozie Version	Components Installed With Oozie
emr-5.36.2	Oozie 5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Topics

Using Oozie with a remote database in Amazon RDS

By default, Oozie user information and query histories are stored in a local MySQL database on the master node. Alternatively, you can create one or more Oozie-enabled clusters using a configuration stored in Amazon S3 and a MySQL database in Amazon Relational Database Service(Amazon RDS). This allows you to persist user information and query history created by Oozie without keeping your Amazon EMR cluster running. We recommend using Amazon S3 server-side encryption to store the configuration file.

First, create the remote database for Oozie.

To create the external MySQL database

1. Open the Amazon RDS console at https://console.aws.amazon.com/rds/.
2. Choose **Launch a DB Instance**.
3. Choose MySQL and then choose **Select**.
4. Leave the default selection of **Multi-AZ Deployment and Provisioned IOPS Storage** and choose **Next**.
5. Leave the Instance Specifications at their defaults, specify Settings, and choose **Next**.
6. On the Configure Advanced Settings page, choose proper security group and database names. The security group you use must at least allow inbound TCP access for port 3306 from the master node of your cluster. If you have not created your cluster at this point, you can allow all hosts to connect to port 3306 and adjust the security group after you have launched the cluster. Choose **Launch DB Instance**.
7. From the RDS Dashboard, select **Instances** and select the instance you have just created. When your database is available, make a note of the dbname, username, password, and RDS instance hostname. You use this information when you create and configure your cluster.

To specify an external MySQL database for Oozie when launching a cluster using the AWS CLI

To specify an external MySQL database for Oozie when launching a cluster using the AWS CLI, use the information you noted when creating your RDS instance for configuring `oozie-site` with a configuration object.

> **Note**
>
> You can create multiple clusters that use the same external database, but each cluster will share query history and user information.

- Using the AWS CLI, create a cluster with Oozie installed, using the external database you created, and referencing a configuration file with a configuration classification for Oozie that specifies the database properties. The following example creates a cluster with Oozie installed, referencing a configuration file in Amazon S3, `myConfig.json`, that specifies the database configuration.

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --release-label emr-7.2.0 --applications Name=Oozie
 Name=Spark Name=Hive \
--instance-type m5.xlarge --instance-count 3 \
--configurations https://s3.amazonaws.com/mybucket/myfolder/myConfig.json --use-
default-roles
```

Example contents of the `myConfig.json` file are shown below. Replace *JDBC URL*, *username*, and *password* with the JDBC URL, user name, and password of your RDS instance.

> **Important**
>
> The JDBC URL must include the database name as a suffix. For example, jdbc:mysql://oozie-external-db.xxxxxxxxxx.us-east-1.rds.amazonaws.com:3306/**dbname**.

```
[{
  "Classification": "oozie-site",
    "Properties": {
        "oozie.service.JPAService.jdbc.driver": "org.mariadb.jdbc.Driver",
        "oozie.service.JPAService.jdbc.url": "JDBC URL",

        "oozie.service.JPAService.jdbc.username": "username",
        "oozie.service.JPAService.jdbc.password": "password"
    },
    "Configurations": []
}]
```

Configure Java version for Oozie

Oozie runs multiple Java Virtual Machine (JVM) processes. This page explains how to configure the Java version for each process.

- **Oozie Server**: Set `JAVA_HOME` in the `oozie-env` classification to update the Java version for the `EmbeddedOozieServer`.
- **Oozie Launcher AM**: *Oozie Launcher AM* is a single-mapper MR job that invokes the appropriate application client libraries such as Hadoop and Hive. Unless otherwise configured, the runtime versions for Oozie Launcher AM are the same as the Java runtimes for Hadoop in the EMR cluster. To configure the Java runtime for Oozie Launcher AM, set the following property in the `workflow.xml` for the job:

```
<property>
     <name>mapred.child.env</name>
     <value>JAVA_HOME=/path/to/JAVA_HOME</value>
 </property>
```

  This property ensures that the Oozie Launcher AM for the Oozie job runs on the Java version that you specify, rather than the Java version that is set in Hadoop.
- **Application Client Executable**: Because Oozie Launcher AM invokes the application client by default, the Java runtime for the client executable is the same as the Oozie Launcher AM.
- **Applications launched by an Oozie job**: Unless otherwise specified, the runtime versions for the actual application JVMs that are launched by an Oozie job are the same as the Java runtimes for Hadoop in the EMR cluster. Depending on the type of Oozie Workflow Action used to launch the application in an Oozie job (Spark or Hive action), you can update the default Java runtime for the actual application JVMs in the `workflow.xml` for the Oozie job.

Oozie release history

The following table lists the version of Oozie included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in Amazon EMR 7.x release versions, Amazon EMR 6.x release versions, or Amazon EMR 5.x release versions.

Oozie version information

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-7.2.0	5.2.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist-

Amazon EMR Release label	Oozie Version	Components installed with Oozie
		cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-5.36.2	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-7.1.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-7.0.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-6.15.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-6.14.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-6.13.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-6.12.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-6.11.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-6.11.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-6.10.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-6.10.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-6.9.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.9.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-6.8.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.8.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-6.7.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.36.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.36.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.6.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.35.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.5.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-6.4.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.3.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-6.3.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.2.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-6.2.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.1.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-6.1.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.0.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-6.0.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.34.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.33.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.33.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.32.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.32.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.31.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.31.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.30.2	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.30.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.30.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.29.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.28.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.28.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.27.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.27.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.26.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.25.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.24.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.24.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.23.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.23.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.22.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.21.2	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.21.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.21.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.20.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.20.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.19.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.19.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.18.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.18.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.17.2	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.17.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.17.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.16.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.16.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.15.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.15.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.14.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.14.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.14.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.13.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.13.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.12.3	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.12.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.12.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.12.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.11.4	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.11.3	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.11.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.11.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.11.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.10.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.10.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.9.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.9.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.8.3	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.8.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.8.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.8.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.7.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.7.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.6.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.6.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.5.4	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.5.3	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.5.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.5.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.5.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.4.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.4.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.3.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.3.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.3.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.2.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.2.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.2.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.2.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.1.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.1.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.0.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.0.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.0.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release label	Oozie Version	Components installed with Oozie
emr-5.0.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Oozie release notes by version

Topics

Amazon EMR 7.1.0 - Oozie changes

Type	Description
Upgrade	Upgrades Netty to 4.1.100.Final.
Upgrade	Upgrades Jetty to 9.4.53.v20231009.

Apache Phoenix

Apache Phoenix is used for OLTP and operational analytics, allowing you to use standard SQL queries and JDBC APIs to work with an Apache HBase backing store. For more information, see Phoenix in 15 minutes or less. Phoenix is included in Amazon EMR release version 4.7.0 and later.

If you upgrade from an earlier version of Amazon EMR to Amazon EMR release version 5.4.0 or later and use secondary indexing, upgrade local indexes as described in the Apache Phoenix documentation. Amazon EMR removes the required configurations from the `hbase-site` classification, but indexes need to be repopulated. Online and offline upgrade of indexes are supported. Online upgrades are the default, which means indexes are repopulated while initializing from Phoenix clients of version 4.8.0 or greater. To specify offline upgrades, set the `phoenix.client.localIndexUpgrade` configuration to false in the `phoenix-site` classification, and then SSH to the master node to run `psql [zookeeper] -1`.

The following table lists the version of Phoenix included in the latest release of the Amazon EMR 7.x series, along with the components that Amazon EMR installs with Phoenix.

For the version of components installed with Phoenix in this release, see Release 7.2.0 Component Versions.

Phoenix version information for emr-7.2.0

Amazon EMR Release Label	Phoenix Version	Components Installed With Phoenix
emr-7.2.0	Phoenix 5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-

Amazon EMR Release Label	Phoenix Version	Components Installed With Phoenix
		region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

The following table lists the version of Phoenix included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with Phoenix.

For the version of components installed with Phoenix in this release, see Release 6.15.0 Component Versions.

Phoenix version information for emr-6.15.0

Amazon EMR Release Label	Phoenix Version	Components Installed With Phoenix
emr-6.15.0	Phoenix 5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

The following table lists the version of Phoenix included in the latest release of the Amazon EMR 5.x series, along with the components that Amazon EMR installs with Phoenix.

For the version of components installed with Phoenix in this release, see Release 5.36.2 Component Versions.

Phoenix version information for emr-5.36.2

Amazon EMR Release Label	Phoenix Version	Components Installed With Phoenix
emr-5.36.2	Phoenix 4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Topics

- Creating a cluster with Phoenix
- Phoenix clients
- Phoenix release history

Creating a cluster with Phoenix

You install Phoenix by choosing the application when you create a cluster in the console or using the AWS CLI. The following procedures and examples show how to create a cluster with Phoenix and HBase. For more information about creating clusters using the console, including **Advanced Options** see Plan and configure clusters in the *Amazon EMR Management Guide*.

To launch a cluster with Phoenix installed using Quick Options for creating a cluster in the console

1. Open the Amazon EMR console at https://console.aws.amazon.com/emr.
2. Choose **Create cluster** to use **Quick Create**.
3. For **Software Configuration**, choose the most recent release appropriate for your application. Phoenix appears as an option only when **Amazon Release Version emr-4.7.0** or later is selected.
4. For **Applications**, choose the second option, **HBase: HBase *ver* with Ganglia *ver*, Hadoop *ver*, Hive *ver*, Hue *ver*, Phoenix *ver*, and ZooKeeper *ver***.
5. Select other options as necessary and then choose **Create cluster**.

> **Note**
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

The following example launches a cluster with Phoenix installed using default configuration settings.

To launch a cluster with Phoenix and HBase using the AWS CLI

- Create the cluster with the following command:

```
aws emr create-cluster --name "Cluster with Phoenix" --release-label emr-7.2.0 \
--applications Name=Phoenix Name=HBase --ec2-attributes KeyName=myKey \
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

Customizing Phoenix configurations

When creating a cluster, you configure Phoenix by setting values in `hbase-site.xml` using the `hbase-site` configuration classification.

For more information, see Configuration and tuning in the Phoenix documentation.

The following example demonstrates using a JSON file stored in Amazon S3 to specify the value of `false` for the `phoenix.schema.dropMetaData` property. Multiple properties can be specified for a single classification. For more information, see Configure applications. The `create-cluster` command then references the JSON file as the `--configurations` parameter.

The contents of the JSON file saved to /mybucket/myfolder/myconfig.json is the following.

```
[
    {
      "Classification": "hbase-site",
      "Properties": {
        "phoenix.schema.dropMetaData": "false"
      }
    }
  ]
```

The `create cluster` command that references the JSON file is shown in the following example.

```
aws emr create-cluster --release-label emr-7.2.0 --applications Name=Phoenix \
Name=HBase --instance-type m5.xlarge --instance-count 2 \
--configurations https://s3.amazonaws.com/mybucket/myfolder/myconfig.json
```

Note

Reconfiguration request for any Phoenix configuration classifications is only supported in Amazon EMR version 5.23.0 and later, and is not supported in Amazon EMR version 5.21.0 or 5.22.0. For more information, see Supplying a configuration for an instance group in a running cluster

Phoenix clients

You connect to Phoenix using either a JDBC client built with full dependencies or using the "thin client" that uses the Phoenix Query Server and can only be run on a master node of a cluster (e.g. by using an SQL client, a step, command line, SSH port forwarding, etc.). When using the "fat" JDBC client, it still needs to have access to all nodes of the cluster because it connects to HBase services directly. The "thin" Phoenix client only needs access to the Phoenix Query Server at a default port 8765. There are several scripts within Phoenix that use these clients.

Use an Amazon EMR step to query using Phoenix

The following procedure restores a snapshot from HBase and uses that data to run a Phoenix query. You can extend this example or create a new script that leverages Phoenix's clients to suit your needs.

1. Create a cluster with Phoenix installed, using the following command:

```
aws emr create-cluster --name "Cluster with Phoenix" --log-uri s3://myBucket/
myLogFolder --release-label emr-7.2.0 \
--applications Name=Phoenix Name=HBase --ec2-attributes KeyName=myKey \
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

2. Create then upload the following files to Amazon S3:

   copySnapshot.sh

```
sudo su hbase -s /bin/sh -c 'hbase snapshot export \
 -D hbase.rootdir=s3://us-east-1.elasticmapreduce.samples/hbase-demo-customer-data/
snapshot/ \
-snapshot customer_snapshot1 \
-copy-to hdfs://masterDNSName:8020/user/hbase \
-mappers 2 -chuser hbase -chmod 700'
```

   runQuery.sh

```
aws s3 cp s3://myBucket/phoenixQuery.sql /home/hadoop/
/usr/lib/phoenix/bin/sqlline-thin.py http://localhost:8765 /home/hadoop/
phoenixQuery.sql
```

   phoenixQuery.sql

Note

You only need to include `COLUMN_ENCODED_BYTES=0` in the following example when you use Amazon EMR versions 5.26.0 and higher.

```
CREATE VIEW "customer" (
pk VARCHAR PRIMARY KEY,
"address"."state" VARCHAR,
"address"."street" VARCHAR,
"address"."city" VARCHAR,
"address"."zip" VARCHAR,
"cc"."number" VARCHAR,
"cc"."expire" VARCHAR,
"cc"."type" VARCHAR,
"contact"."phone" VARCHAR)
COLUMN_ENCODED_BYTES=0;

CREATE INDEX my_index ON "customer" ("customer"."state") INCLUDE("PK",
 "customer"."city", "customer"."expire", "customer"."type");

SELECT "customer"."type" AS credit_card_type, count(*) AS num_customers FROM
 "customer" WHERE "customer"."state" = 'CA' GROUP BY "customer"."type";
```

Use the AWS CLI to submit the files to the S3 bucket:

```
aws s3 cp copySnapshot.sh s3://myBucket/
aws s3 cp runQuery.sh s3://myBucket/
aws s3 cp phoenixQuery.sql s3://myBucket/
```

3. Create a table using the following step submitted to the cluster that you created in Step 1:

   createTable.json

```
[
  {
    "Name": "Create HBase Table",
    "Args": ["bash", "-c", "echo $'create \"customer\",\"address\",\"cc\",\"contact
\"' | hbase shell"],
    "Jar": "command-runner.jar",
```

```
    "ActionOnFailure": "CONTINUE",
    "Type": "CUSTOM_JAR"
  }
]
```

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps file://./createTable.json
```

4. Use `script-runner.jar` to run the `copySnapshot.sh` script that you previously uploaded to your S3 bucket:

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps Type=CUSTOM_JAR,Name="HBase Copy Snapshot",ActionOnFailure=CONTINUE,\
Jar=s3://region.elasticmapreduce/libs/script-runner/script-
runner.jar,Args=["s3://myBucket/copySnapshot.sh"]
```

   This runs a MapReduce job to copy your snapshot data to the cluster HDFS.

5. Restore the snapshot that you copied to the cluster using the following step:

   restoreSnapshot.json

```
[
  {
    "Name": "restore",
    "Args": ["bash", "-c", "echo $'disable \"customer\"; restore_snapshot
 \"customer_snapshot1\"; enable \"customer\"' | hbase shell"],
    "Jar": "command-runner.jar",
    "ActionOnFailure": "CONTINUE",
    "Type": "CUSTOM_JAR"
  }
]
```

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps file://./restoreSnapshot.json
```

6. Use `script-runner.jar` to run the `runQuery.sh` script that you previously uploaded to your S3 bucket:

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps Type=CUSTOM_JAR,Name="Phoenix Run Query",ActionOnFailure=CONTINUE,\
```

```
Jar=s3://region.elasticmapreduce/libs/script-runner/script-
runner.jar,Args=["s3://myBucket/runQuery.sh"]
```

The query runs and returns the results to the step's `stdout`. It may take a few minutes for this step to complete.

7. Inspect the results of the step's `stdout` at the log URI that you used when you created the cluster in Step 1. The results should look like the following:

```
+------------------------------------------+-----------------------------------+
|             CREDIT_CARD_TYPE             |           NUM_CUSTOMERS           |
+------------------------------------------+-----------------------------------+
| american_express                         | 5728                              |
| dankort                                  | 5782                              |
| diners_club                              | 5795                              |
| discover                                 | 5715                              |
| forbrugsforeningen                       | 5691                              |
| jcb                                      | 5762                              |
| laser                                    | 5769                              |
| maestro                                  | 5816                              |
| mastercard                               | 5697                              |
| solo                                     | 5586                              |
| switch                                   | 5781                              |
| visa                                     | 5659                              |
+------------------------------------------+-----------------------------------+
```

Phoenix release history

The following table lists the version of Phoenix included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in Amazon EMR 7.x release versions, Amazon EMR 6.x release versions, or Amazon EMR 5.x release versions.

Phoenix version information

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-7.2.0	5.1.3	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist-

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
		cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.36.2	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-7.1.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-7.0.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-6.15.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-6.14.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-6.13.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-6.12.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-6.11.1	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-6.11.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-6.10.1	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-6.10.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-6.9.1	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-6.9.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-6.8.1	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-6.8.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-6.7.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.36.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.36.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-6.6.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.35.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.5.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-6.4.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.3.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-6.3.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.2.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-6.2.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.1.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-6.1.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.0.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-6.0.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.34.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.33.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.33.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.32.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.32.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.31.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.31.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.30.2	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.30.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.30.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.29.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.28.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.28.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.27.1	4.14.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.27.0	4.14.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.26.0	4.14.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.25.0	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.24.1	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.24.0	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.23.1	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.23.0	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.22.0	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.21.2	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.21.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.21.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.20.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.20.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.19.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.19.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.18.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.18.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.17.2	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.17.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.17.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.16.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.16.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.15.1	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.15.0	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.14.2	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.14.1	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.14.0	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.13.1	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.13.0	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.12.3	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.12.2	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.12.1	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.12.0	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.11.4	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.11.3	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.11.2	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.11.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.11.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.10.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.10.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.9.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.9.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.8.3	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.8.2	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.8.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.8.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.7.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.7.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.6.1	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.6.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.5.4	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.5.3	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.5.2	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.5.1	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.5.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.4.1	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.4.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.3.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.3.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.3.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.2.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.2.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.2.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.2.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.1.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.1.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.0.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.0.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.0.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-5.0.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.9.6	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-4.9.5	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.9.4	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-4.9.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.9.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-4.9.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.8.5	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-4.8.4	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.8.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-4.8.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.8.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-4.8.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.7.4	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-4.7.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.7.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	Phoenix Version	Components installed with Phoenix
emr-4.7.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.7.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Apache Pig

Apache Pig is an open-source Apache library that runs on top of Hadoop, providing a scripting language that you can use to transform large data sets without having to write complex code in a lower level computer language like Java. The library takes SQL-like commands written in a language called Pig Latin and converts those commands into Tez jobs based on directed acyclic graphs (DAGs) or MapReduce programs. Pig works with structured and unstructured data in a variety of formats. For more information about Pig, see http://pig.apache.org/.

You can execute Pig commands interactively or in batch mode. To use Pig interactively, create an SSH connection to the master node and submit commands using the Grunt shell. To use Pig in batch mode, write your Pig scripts, upload them to Amazon S3, and submit them as cluster steps. For more information on submitting work to a cluster, see Submit work to a cluster in the *Amazon EMR Management Guide*.

When you use Pig to write output to an HCatalog table in Amazon S3, disable Amazon EMR direct write by setting the `mapred.output.direct.NativeS3FileSystem` and `mapred.output.direct.EmrFileSystem` properties to `false`. For more information, see Using HCatalog. Within a Pig script, you can use the `SET mapred.output.direct.NativeS3FileSystem false` and `SET mapred.output.direct.EmrFileSystem false` commands.

The following table lists the version of Pig included in the latest release of the Amazon EMR 7.x series, along with the components that Amazon EMR installs with Pig.

For the version of components installed with Pig in this release, see Release 7.2.0 Component Versions.

Pig version information for emr-7.2.0

Amazon EMR Release Label	Pig Version	Components Installed With Pig
emr-7.2.0	Pig 0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-

Amazon EMR Release Label	Pig Version	Components Installed With Pig
		hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

The following table lists the version of Pig included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with Pig.

For the version of components installed with Pig in this release, see Release 6.15.0 Component Versions.

Pig version information for emr-6.15.0

Amazon EMR Release Label	Pig Version	Components Installed With Pig
emr-6.15.0	Pig 0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

The following table lists the version of Pig included in the latest release of the Amazon EMR 5.x series, along with the components that Amazon EMR installs with Pig.

For the version of components installed with Pig in this release, see Release 5.36.2 Component Versions.

Pig version information for emr-5.36.2

Amazon EMR Release Label	Pig Version	Components Installed With Pig
emr-5.36.2	Pig 0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Topics

Submit Pig work

This section demonstrates submitting Pig work to an Amazon EMR cluster. The examples that follow generate a report containing the total bytes transferred, a list of the top 50 IP addresses, a list of the top 50 external referrers, and the top 50 search terms using Bing and Google. The Pig script is located in the Amazon S3 bucket `s3://elasticmapreduce/samples/pig-apache/do-reports2.pig`. Input data is located in the Amazon S3 bucket `s3://elasticmapreduce/samples/pig-apache/input`. The output is saved to an Amazon S3 bucket.

Submit Pig work using the Amazon EMR console

This example describes how to use the Amazon EMR console to add a Pig step to a cluster.

To submit a Pig step

1. Open the Amazon EMR console at https://console.aws.amazon.com/emr.
2. Choose **Create cluster** to create a cluster with Pig installed. For steps on how to create a cluster, see Plan and configure an Amazon EMR cluster.
3. Open a terminal and SSH into the master node of your cluster following the steps outlined in Connect to the master node using SSH. Once you've done that, run the following steps.

```
sudo mkdir -p /home/hadoop/lib/pig/
sudo aws s3 cp s3://elasticmapreduce/libs/pig/0.3/piggybank-0.3-amzn.jar /home/
hadoop/lib/pig/piggybank.jar
```

4. In the console, click **Cluster List** and select the name of the cluster you created.
5. Scroll to the **Steps** section and expand it, then choose **Add step**.
6. In the **Add Step** dialog:
   - For **Step type**, choose **Pig program**.
   - For **Name**, accept the default name (Pig program) or type a new name.
   - For **Script S3 location**, type the location of the Pig script. For example: **`s3://elasticmapreduce/samples/pig-apache/do-reports2.pig`**.
   - For **Input S3 location**, type the location of the input data. For example: **`s3://elasticmapreduce/samples/pig-apache/input`**.
   - For **Output S3 location**, type or browse to the name of your Amazon S3 output bucket.
   - For **Arguments**, leave the field blank.
   - For **Action on failure**, accept the default option (**Continue**).
7. Choose **Add**. The step appears in the console with a status of Pending.
8. The status of the step changes from Pending to Running to Completed as the step runs. To update the status, choose the **Refresh** icon above the **Actions** column. When your step is complete, check your Amazon S3 bucket to confirm your Pig step's output files are there.

Submit Pig work using the AWS CLI

To submit a Pig step using the AWS CLI

When you launch a cluster using the AWS CLI, use the `--applications` parameter to install Pig. To submit a Pig step, use the `--steps` parameter.

1. To launch a cluster with Pig installed, type the following command, replacing *myKey* and *DOC-EXAMPLE-BUCKET/* with the name of your EC2 key pair and Amazon S3 bucket.

```
aws emr create-cluster \
--name "Test cluster" \
--log-uri s3://DOC-EXAMPLE-BUCKET/ \
--release-label emr-5.36.2 \
--applications Name=Pig \
--use-default-roles \
--ec2-attributes KeyName=myKey \
--instance-type m5.xlarge \
--instance-count 3
```

Note

Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

When you specify the instance count without using the `--instance-groups` parameter, a single master node is launched, and the remaining instances are launched as core nodes. All nodes use the instance type specified in the command.

Note

If you have not previously created the default EMR service role and EC2 instance profile, type `aws emr create-default-roles` to create them before typing the `create-cluster` subcommand.

2. To submit a Pig step, enter the following command, replacing *myClusterId* and *DOC-EXAMPLE-BUCKET* with your cluster ID and name of your Amazon S3 bucket.

```
aws emr add-steps \
--cluster-id myClusterId \
--steps Type=PIG,Name="Pig Program",ActionOnFailure=CONTINUE,Args=[-f,s3://
elasticmapreduce/samples/pig-apache/do-reports2.pig,-p,INPUT=s3://elasticmapreduce/
samples/pig-apache/input,-p,OUTPUT=s3://DOC-EXAMPLE-BUCKET/pig-apache/output]
```

This command will return a step ID, which you can use to check the `State` of your step.

3. Query the status of your step with the `describe-step` command.

```
aws emr describe-step --cluster-id myClusterId --step-id s-1XXXXXXXXXXA
```

The `State` of the step changes from `PENDING` to `RUNNING` to `COMPLETED` as the step runs. When your step is complete, check your Amazon S3 bucket to confirm your Pig step's output files are there.

For more information about using Amazon EMR commands in the AWS CLI, see the AWS CLI Command Reference.

Call user-defined functions from Pig

Pig provides the ability to call user-defined functions (UDFs) from within Pig scripts. You can do this to implement custom processing to use in your Pig scripts. The languages currently supported are Java, Python/Jython, and JavaScript (though JavaScript support is still experimental.)

The following sections describe how to register your functions with Pig so you can call them either from the Pig shell or from within Pig scripts. For more information about using UDFs with Pig, see Pig documentation for your version of Pig.

Call JAR files from Pig

You can use custom JAR files with Pig using the `REGISTER` command in your Pig script. The JAR file is local or a remote file system such as Amazon S3. When the Pig script runs, Amazon EMR downloads the JAR file automatically to the master node and then uploads the JAR file to the Hadoop distributed cache. In this way, the JAR file is automatically used as necessary by all instances in the cluster.

To use JAR files with Pig

1. Upload your custom JAR file into Amazon S3.
2. Use the REGISTER command in your Pig script to specify the bucket on Amazon S3 of the custom JAR file.

```
REGISTER s3://mybucket/path/mycustomjar.jar;
```

Call Python/Jython scripts from Pig

You can register Python scripts with Pig and then call functions in those scripts from the Pig shell or in a Pig script. You do this by specifying the location of the script with the register keyword.

Because Pig in written in Java, it uses the Jython script engine to parse Python scripts. For more information about Jython, go to http://www.jython.org/.

To call a Python/Jython script from Pig

1. Write a Python script and upload the script to a location in Amazon S3. This should be a bucket owned by the same account that creates the Pig cluster, or that has permissions set so the account that created the cluster can access it. In this example, the script is uploaded to s3://mybucket/pig/python.
2. Start a Pig cluster. If you are accessing Pig from the Grunt shell, run an interactive cluster. If you are running Pig commands from a script, start a scripted Pig cluster. This example starts an interactive cluster. For more information about how to create a Pig cluster, see Submit Pig work.
3. For an interactive cluster, use SSH to connect into the master node and run the Grunt shell. For more information, see SSH into the master node.
4. Run the Grunt shell for Pig by typing pig at the command line:

```
pig
```

5. Register the Jython library and your Python script with Pig using the register keyword at the Grunt command prompt, as shown in the following command, where you would specify the location of your script in Amazon S3:

```
grunt> register 'lib/jython.jar';
```

```
grunt> register 's3://mybucket/pig/python/myscript.py' using jython as myfunctions;
```

6. Load the input data. The following example loads input from an Amazon S3 location:

```
grunt> input = load 's3://mybucket/input/data.txt' using TextLoader as
 (line:chararray);
```

7. You can now call functions in your script from within Pig by referencing them using `myfunctions`:

```
grunt> output=foreach input generate myfunctions.myfunction($1);
```

Pig release history

The following table lists the version of Pig included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in Amazon EMR 7.x release versions, Amazon EMR 6.x release versions, or Amazon EMR 5.x release versions.

Pig version information

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-7.2.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.36.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-7.1.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-7.0.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker
emr-6.15.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-6.14.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker
emr-6.13.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-6.12.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker
emr-6.11.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-6.11.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker
emr-6.10.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-6.10.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker
emr-6.9.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-6.9.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-6.8.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-6.8.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-6.7.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.36.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.36.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-6.6.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.35.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-6.5.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-6.4.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-6.3.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-6.3.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-6.2.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-6.2.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-6.1.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-6.1.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.34.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.33.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.33.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.32.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.32.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.31.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.31.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.30.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.30.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.30.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.29.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.28.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.28.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.27.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.27.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.26.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.25.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.24.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.24.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.23.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.23.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.22.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.21.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.21.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.21.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.20.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.20.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.19.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.19.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.18.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.18.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.17.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.17.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.17.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.16.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.16.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.15.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.15.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.14.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.14.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.14.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.13.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.13.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.12.3	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.12.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.12.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.12.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.11.4	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.11.3	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.11.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.11.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.11.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.10.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.10.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.9.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.9.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.8.3	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.8.2	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.8.1	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.8.0	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.7.1	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.7.0	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.6.1	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.6.0	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.5.4	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.5.3	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.5.2	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.5.1	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.5.0	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.4.1	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.4.0	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.3.2	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.3.1	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.3.0	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.2.3	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.2.2	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.2.1	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.2.0	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.1.1	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.1.0	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.0.3	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.0.2	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-5.0.1	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.0.0	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-4.9.6	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.9.5	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.9.4	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-4.9.3	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.9.2	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.9.1	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-4.8.5	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.8.4	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.8.3	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-4.8.2	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.8.1	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.8.0	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-4.7.4	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.7.3	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.7.2	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-4.7.1	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.7.0	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.6.1	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-4.6.0	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.5.0	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.4.0	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-4.3.0	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.2.0	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.1.0	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Amazon EMR Release label	Pig Version	Components installed with Pig
emr-4.0.0	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Presto and Trino

Note

PrestoSQL was renamed to Trino in December 2020. Amazon EMR versions 6.4.0 and later use the name Trino, while earlier release versions use the name PrestoSQL.

Presto is a fast SQL query engine designed for interactive analytic queries over large datasets from multiple sources. For more information, see the Presto website. Presto is included in Amazon EMR releases 5.0.0 and later. Earlier releases include Presto as a sandbox application. For more information, see Amazon EMR 4.x release versions. Amazon EMR release versions 6.1.0 and later support Trino (PrestoSQL) in addition to Presto. For more information, see PrestoDB and Trino installation.

The following table lists the version of Presto included in the latest release of the Amazon EMR 7.x series, along with the components that Amazon EMR installs with Presto.

For the version of components installed with Presto in this release, see Release 7.2.0 Component Versions.

Presto version information for emr-7.2.0

Amazon EMR Release Label	Presto Version	Components Installed With Presto
emr-7.2.0	Presto 0.285	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

The following table lists the version of Presto included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with Presto.

For the version of components installed with Presto in this release, see Release 6.15.0 Component Versions.

Presto version information for emr-6.15.0

Amazon EMR Release Label	Presto Version	Components Installed With Presto
emr-6.15.0	Presto 0.283	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

The following table lists the version of Presto included in the latest release of the Amazon EMR 5.x series, along with the components that Amazon EMR installs with Presto.

For the version of components installed with Presto in this release, see Release 5.36.2 Component Versions.

Presto version information for emr-5.36.2

Amazon EMR Release Label	Presto Version	Components Installed With Presto
emr-5.36.2	Presto 0.267	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode,

Amazon EMR Release Label	Presto Version	Components Installed With Presto
		hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server , hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

The following table lists the version of Trino (PrestoSQL) included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with Trino (PrestoSQL).

For the version of components installed with Trino (PrestoSQL) in this release, see Release 6.15.0 Component Versions.

Trino (PrestoSQL) version information for emr-6.15.0

Amazon EMR Release Label	Trino (PrestoSQL) Version	Components Installed With Trino (PrestoSQL)
emr-6.15.0	Trino (PrestoSQL) 426	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server , hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Topics

Using Presto with the AWS Glue Data Catalog

Using Amazon EMR release version 5.10.0 and later, you can specify the AWS Glue Data Catalog as the default Hive metastore for Presto. We recommend this configuration when you require a persistent metastore or a metastore shared by different clusters, services, applications, or AWS accounts.

AWS Glue is a fully managed extract, transform, and load (ETL) service that makes it simple and cost-effective to categorize your data, clean it, enrich it, and move it reliably between various data stores. The AWS Glue Data Catalog provides a unified metadata repository across a variety of data sources and data formats, integrating with Amazon EMR as well as Amazon RDS, Amazon Redshift, Redshift Spectrum, Athena, and any application compatible with the Apache Hive metastore. AWS Glue crawlers can automatically infer schema from source data in Amazon S3 and store the associated metadata in the Data Catalog. For more information about the Data Catalog, see Populating the AWS Glue Data Catalog in the *AWS Glue Developer Guide*.

Separate charges apply for AWS Glue. There is a monthly rate for storing and accessing the metadata in the Data Catalog, an hourly rate billed per minute for AWS Glue ETL jobs and crawler runtime, and an hourly rate billed per minute for each provisioned development endpoint. The Data Catalog allows you to store up to a million objects at no charge. If you store more than a million objects, you are charged USD$1 for each 100,000 objects over a million. An object in the Data Catalog is a table, partition, or database. For more information, see Glue Pricing.

Important

If you created tables using Amazon Athena or Amazon Redshift Spectrum before August 14, 2017, databases and tables are stored in an Athena-managed catalog, which is separate from the AWS Glue Data Catalog. To integrate Amazon EMR with these tables, you must upgrade to the AWS Glue Data Catalog. For more information, see Upgrading to the AWS Glue Data Catalog in the *Amazon Athena User Guide*.

Specifying AWS Glue Data Catalog as the metastore

You can specify the AWS Glue Data Catalog as the metastore using the AWS Management Console, AWS CLI, or Amazon EMR API. When you use the CLI or API, you use the configuration classification for Presto to specify the Data Catalog. In addition, with Amazon EMR 5.16.0 and later, you can use the configuration classification to specify a Data Catalog in a different AWS account. When you use the console, you can specify the Data Catalog using **Advanced Options** or **Quick Options**.

Console

To specify AWS Glue Data Catalog as the Hive metastore with the new console

1. Sign in to the AWS Management Console, and open the Amazon EMR console at https://console.aws.amazon.com/emr.
2. Under **EMR on EC2** in the left navigation pane, choose **Clusters**, and then choose **Create cluster**.
3. Under **Application bundle**, choose **Presto**.
4. Under **AWS Glue Data Catalog settings**, select the **Use for Presto table metadata** check box.
5. Choose any other options that apply to your cluster.
6. To launch your cluster, choose **Create cluster**.

CLI

To specify the AWS Glue Data Catalog as the default Hive metastore using the AWS CLI

For examples of how to specify the following configuration classifications when you create a cluster, see Configure applications.

Amazon EMR 5.16.0 and later

- Set the `hive.metastore` property to `glue` as shown in the following JSON example.

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.metastore": "glue"
    }
  }
]
```

To specify a Data Catalog in a different AWS account, add the `hive.metastore.glue.catalogid` property as shown in the following JSON example. Replace *`acct-id`* with the AWS account of the Data Catalog. Using a Data Catalog in another AWS account is not available using Amazon EMR version 5.15.0 and earlier.

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.metastore": "glue",
      "hive.metastore.glue.catalogid": "acct-id"
    }
  }
]
```

Amazon EMR 5.10.0 through 5.15.0

Set the `hive.metastore.glue.datacatalog.enabled` property to `true`, as shown in the following JSON example:

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.metastore.glue.datacatalog.enabled": "true"
    }
  }
]
```

Amazon EMR 6.1.0 and later using PrestoSQL (Trino)

Starting with EMR version 6.1.0, PrestoSQL also supports Glue as the default Hive metastore. Use the `prestosql-connector-hive` configuration classification and set the `hive.metastore` property to `glue`, as shown in the following JSON example.

Amazon EMR versions 6.4.0 and later use the new name Trino instead of PrestoSQL. If you use Trino, replace *`prestosql-connector-hive`* in the following configuration classification with `trino-connector-hive`.

```
[
  {
    "Classification": "prestosql-connector-hive",
    "Properties": {
      "hive.metastore": "glue"
    }
  }
]
```

To switch metastores on a long-running cluster, you can manually set these values as appropriate for your release version by connecting to the master node, editing the property values in the `/etc/presto/conf/catalog/hive.properties` file directly, and restarting the Presto server (`sudo restart presto-server`). If you use this method with Amazon EMR 5.15.0 and earlier, make sure that `hive.table-statistics-enabled` is set to `false`. This setting is not required when using release versions 5.16.0 and later; nevertheless, table and partition statistics are not supported.

IAM permissions

The EC2 instance profile for a cluster must have IAM permissions for AWS Glue actions. In addition, if you enable encryption for AWS Glue Data Catalog objects, the role must also be allowed to encrypt, decrypt and generate the AWS KMS key used for encryption.

Permissions for AWS Glue actions

If you use the default EC2 instance profile for Amazon EMR, no action is required. The `AmazonElasticMapReduceforEC2Role` managed policy that is attached to the

EMR_EC2_DefaultRole allows all necessary AWS Glue actions. However, if you specify a custom EC2 instance profile and permissions, you must configure the appropriate AWS Glue actions. Use the AmazonElasticMapReduceforEC2Role managed policy as a starting point. For more information, see Service role for cluster EC2 instances (EC2 instance profile) in the *Amazon EMR Management Guide*.

Permissions for encrypting and decrypting AWS Glue Data Catalog

Your instance profile needs permission to encrypt and decrypt data using your key. You do *not* need to configure these permissions if both of the following statements apply:

- You enable encryption for AWS Glue Data Catalog objects using managed keys for AWS Glue.
- You use a cluster that's in the same AWS account as the AWS Glue Data Catalog.

Otherwise, you must add the following statement to the permissions policy attached to your EC2 instance profile.

```
[
    {
        "Version": "2012-10-17",
        "Statement": [
            {
                "Effect": "Allow",
                "Action": [
                    "kms:Decrypt",
                    "kms:Encrypt",
                    "kms:GenerateDataKey"
                ],
                "Resource": "arn:aws:kms:region:acct-
id:key/12345678-1234-1234-1234-123456789012"
            }
        ]
    }
]
```

For more information about AWS Glue Data Catalog encryption, see Encrypting your data catalog in the *AWS Glue Developer Guide*.

Resource-based permissions

If you use AWS Glue in conjunction with Hive, Spark, or Presto in Amazon EMR, AWS Glue supports resource-based policies to control access to Data Catalog resources. These resources include databases, tables, connections, and user-defined functions. For more information, see AWS Glue Resource Policies in the *AWS Glue Developer Guide*.

When using resource-based policies to limit access to AWS Glue from within Amazon EMR, the principal that you specify in the permissions policy must be the role ARN associated with the EC2 instance profile that is specified when a cluster is created. For example, for a resource-based policy attached to a catalog, you can specify the role ARN for the default service role for cluster EC2 instances, *EMR_EC2_DefaultRole* as the `Principal`, using the format shown in the following example:

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

The *acct-id* can be different from the AWS Glue account ID. This enables access from EMR clusters in different accounts. You can specify multiple principals, each from a different account.

Considerations when using AWS Glue Data Catalog

Consider the following items when using AWS Glue Data Catalog as a metastore with Presto:

- Renaming tables from within AWS Glue is not supported.
- When you create a Hive table without specifying a `LOCATION`, the table data is stored in the location specified by the `hive.metastore.warehouse.dir` property. By default, this is a location in HDFS. If another cluster needs to access the table, it fails unless it has adequate permissions to the cluster that created the table. Furthermore, because HDFS storage is transient, if the cluster terminates, the table data is lost, and the table must be recreated. We recommend that you specify a `LOCATION` in Amazon S3 when you create a Hive table using AWS Glue. Alternatively, you can use the `hive-site` configuration classification to specify a location in Amazon S3 for `hive.metastore.warehouse.dir`, which applies to all Hive tables. If a table is created in an HDFS location and the cluster that created it is still running, you can update the table location to Amazon S3 from within AWS Glue. For more information, see Working with Tables on the AWS Glue Console in the *AWS Glue Developer Guide*.
- Partition values containing quotes and apostrophes are not supported, for example, `PARTITION (owner="Doe's")`.
- Column statistics are supported for emr-5.31.0 and later.

- Using Hive authorization is not supported. As an alternative, consider using AWS Glue Resource-Based Policies. For more information, see Use Resource-Based Policies for Amazon EMR Access to AWS Glue Data Catalog.

Using S3 Select Pushdown with Presto to improve performance

With Amazon EMR release version 5.18.0 and later, you can use S3 select Pushdown with Presto on Amazon EMR. This feature allows Presto to "push down" the computational work of projection operations (for example, `SELECT`) and predicate operations (for example, `WHERE`) to Amazon S3. This allows queries to retrieve only required data from Amazon S3, which can improve performance and reduce the amount of data transferred between Amazon EMR and Amazon S3 in some applications.

Is S3 Select Pushdown right for my application?

We recommend that you benchmark your applications with and without S3 Select Pushdown to see if using it may be suitable for your application.

Use the following guidelines to determine if your application is a candidate for using S3 Select:

- Your query filters out more than half of the original data set.
- Your query filter predicates use columns that have a data type supported by Presto and S3 Select. The timestamp, real, and double data types are not supported by S3 Select Pushdown. We recommend using the decimal data type for numerical data. For more information about supported data types for S3 Select, see Data types in the *Amazon Simple Storage Service User Guide*.
- Your network connection between Amazon S3 and the Amazon EMR cluster has good transfer speed and available bandwidth. Amazon S3 does not compress HTTP responses, so the response size is likely to increase for compressed input files.

Considerations and limitations

- Only objects stored in CSV format are supported. Objects can be uncompressed or optionally compressed with gzip or bzip2.
- The `AllowQuotedRecordDelimiters` property is not supported. If this property is specified, the query fails.

- Amazon S3 server-side encryption with customer-provided encryption keys (SSE-C) and client-side encryption are not supported.
- S3 Select Pushdown is not a substitute for using columnar or compressed file formats such as ORC or Parquet.

Enabling S3 Select Pushdown with PrestoDB or Trino

To enable S3 Select Pushdown for PrestoDB on Amazon EMR, use the `presto-connector-hive` configuration classification to set `hive.s3select-pushdown.enabled` to `true` as shown in the example below. For more information, see Configure applications. The hive.s3select-pushdown.max-connections value must also be set. For most applications, the default setting of *500* should be adequate. For more information, see Understanding and tuning hive.s3select-pushdown.max-connections below.

For PrestoSQL on EMR versions 6.1.0 - 6.3.0, replace *presto-connector-hive* in the example below with `prestosql-connector-hive`.

Amazon EMR versions 6.4.0 and later use the new name Trino instead of PrestoSQL. If you use Trino, replace *presto-connector-hive* in the example below with `trino-connector-hive`

```
[
    {
        "classification": "presto-connector-hive",
        "properties": {
            "hive.s3select-pushdown.enabled": "true",
            "hive.s3select-pushdown.max-connections": "500"
        }
    }
]
```

Understanding and tuning hive.s3select-pushdown.max-connections

By default, Presto uses EMRFS as its file system. The setting `fs.s3.maxConnections` in the `emrfs-site` configuration classification specifies the maximum allowable client connections to Amazon S3 through EMRFS for Presto. By default, this is 500. S3 Select Pushdown bypasses EMRFS when accessing Amazon S3 for predicate operations. In this case, the value of `hive.s3select-pushdown.max-connections` determines the maximum number of client connections allowed for those operations from worker nodes. However, any requests to Amazon S3 that Presto initiates that are not pushed down—for example, GET operations—continue to be governed by the value of `fs.s3.maxConnections`.

If your application experiences the error "Timeout waiting for connection from pool," increase the value of both `hive.s3select-pushdown.max-connections` and `fs.s3.maxConnections`.

Adding database connectors

You can use configuration classifications to configure JDBC connector properties when you create a cluster. Configuration classifications begin with `presto-connector`, for example, `presto-connector-postgresql`. The available configuration classifications depend on the Amazon EMR release version. For the configuration classifications available with the most recent release version, see the section called "5.36.2 configuration classifications" for Amazon EMR 5.36.2. If you are using a different version of Amazon EMR, see Amazon EMR 5.x release versions for the configuration classifications. For more information about the properties that can be configured with each connector, see https://prestodb.io/docs/current/connector.html.

Example —configuring a cluster with the PostgreSQL JDBC connector

To launch a cluster with the PostgreSQL connector installed and configured, first create a JSON file that specifies the configuration classification—for example, `myConfig.json`—with the following content, and save it locally.

Replace the connection properties as appropriate for your setup and as shown in the PostgreSQL connector topic in Presto Documentation.

```
[
  {
    "Classification": "presto-connector-postgresql",
    "Properties": {
      "connection-url": "jdbc:postgresql://example.net:5432/database",
      "connection-user": "MYUSER",
      "connection-password": "MYPASS"
    },
    "Configurations": []
  }
]
```

When you create the cluster, reference the path to the JSON file using the `--configurations` option as shown in the following example, where `myConfig.json` is in the same directory where you run the command:

```
aws emr create-cluster --name PrestoConnector --release-label emr-5.36.2 --instance-type m5.xlarge \
--instance-count 2 --applications Name=Hadoop Name=Hive Name=Pig Name=Presto \
--use-default-roles --ec2-attributes KeyName=myKey \
--log-uri s3://my-bucket/logs --enable-debugging \
--configurations file://myConfig.json
```

Using SSL/TLS and configuring LDAPS with Presto on Amazon EMR

With Amazon EMR release version 5.6.0 and later, you can enable SSL/TLS to help secure internal communication between Presto nodes. You do this by setting up a security configuration for in-transit encryption. For more information, see Encryption options and Use security configurations to set up cluster security in the *Amazon EMR Management Guide*.

When you use a security configuration with in-transit encryption, Amazon EMR does the following for Presto:

- Distributes the encryption artifacts, or certificates, that you specify for in-transit encryption throughout the Presto cluster. For more information, see Providing certificates for in-transit data encryption.
- Sets the following properties using the `presto-config` configuration classification, which corresponds to the `config.properties` file for Presto:
  - Sets `http-server.http.enabled` to `false` on all nodes, which disables HTTP in favor of HTTPS. This requires you to provide certificates that work for public and private DNS when setting up the security configuration for in-transit encryption. One way to do this is to use SAN (Subject Alternative Name) certificates which support multiple domains.
  - Sets `http-server.https.*` values. For configuration details, see LDAP authentication in Presto documentation.
- For PrestoSQL (Trino) on EMR version 6.1.0 and later, Amazon EMR automatically configures a shared secret key for secure internal communication between cluster nodes. You don't need to do any additional configuration to enable this security feature, and you can override the

configuration with your own secret key. For information about Trino internal authentication, see Trino 353 documentation: Secure internal communication.

In addition, with Amazon EMR release version 5.10.0 and later, you can set up LDAP authentication for client connections to the Presto coordinator using HTTPS. This setup uses secure LDAP (LDAPS). TLS must be enabled on your LDAP server, and the Presto cluster must use a security configuration with in-transit data encryption enabled. Additional configuration is required. The configuration options are different depending on the release version of Amazon EMR that you use. For more information, see Using LDAP authentication for Presto on Amazon EMR.

Presto on Amazon EMR uses port 8446 for internal HTTPS by default. The port used for internal communication must be the same port used for client HTTPS access to the Presto coordinator. The `http-server.https.port` property in the `presto-config` configuration classification specifies the port.

Using LDAP authentication for Presto on Amazon EMR

Follow the steps in this section to configure LDAP. See each step for examples and links to more information.

Steps to Configure LDAP Authentication

- Step 1: Gather information about your LDAP server and copy the server certificate to Amazon S3
- Step 2: Set up a security configuration
- Step 3: Create a configuration JSON with Presto properties for LDAP
- Step 4: Create the script to copy the LDAP server certificate and upload it to Amazon S3
- Step 5: Create the cluster

Step 1: Gather information about your LDAP server and copy the server certificate to Amazon S3

You'll need the information and items in the following section from your LDAP server to configure LDAP authentication.

The IP address or host name of the LDAP server

The Presto coordinator on the Amazon EMR master node must be able to reach the LDAP server at the specified IP address or host name. By default, Presto communicates with the LDAP server using

LDAPS over port 636. If your LDAP implementation requires a custom port, you can specify it using the `ldap.url` property with Amazon EMR 5.16.0 or later, or using `authentication.ldap.url` with earlier versions. Substitute the custom port for 636 as shown in the `presto-config` configuration classification examples in Step 3: Create a configuration JSON with Presto properties for LDAP. Ensure that any firewalls and security groups allow inbound and outbound traffic on port 636 (or your custom port) and also port 8446 (or your custom port), which is used for internal cluster communications.

The LDAP server certificate

You must upload the certificate file to a secure location in Amazon S3. For more information, see How do I upload files and folders to an S3 Bucket in the *Amazon Simple Storage Service User Guide*. You create a bootstrap action that copies this certificate from Amazon S3 to each node in the cluster when the cluster launches. In Step 4: Create the script to copy the LDAP server certificate and upload it to Amazon S3. The example certificate is *`s3://MyBucket/ldap_server.crt`*.

The LDAP server's settings for anonymous binding

If anonymous binding is disabled on PrestoDB, you need the user ID (UID) and password of an account with permissions to bind to the LDAP server so that the PrestoDB server can establish a connection. You specify the UID and password using the `internal-communication.authentication.ldap.user` and `internal-communication.authentication.ldap.password` properties in the `presto-config` configuration classification. Amazon EMR 5.10.0 does not support these settings, so anonymous binding must be supported on the LDAP server when you use this release version.

Note that Trino doesn't require the anonymous binding configuration.

To get the status of anonymous binding on the LDAP server

- Use the ldapwhoami command from a Linux client, as shown in the following example:

  ```
  ldapwhoami -x -H ldaps://LDAPServerHostNameOrIPAddress
  ```

  If anonymous binding is not allowed, the command returns the following:

  ```
  ldap_bind: Inappropriate authentication (48)
  additional info: anonymous bind disallowed
  ```

To verify that an account has permissions to an LDAP server that uses simple authentication

- Use theldapwhoami command from a Linux client, as shown in the following example. The example uses a fictitious user, *presto*, stored in an Open LDAP server running on an EC2 instance with the fictitious host name *ip-xxx-xxx-xxx-xxx.ec2.internal*. The user is associated with the organizational unit (OU) *admins* and with the password *123456*:

  ```
  ldapwhoami -x -w "123456" -D uid=presto,ou=admins,dc=ec2,dc=internal -H ldaps://ip-xxx-xxx-xxx-xxx.ec2.internal
  ```

  If the account is valid and has appropriate permissions, the command returns:

  ```
  dn:uid=presto,ou=admins,dc=ec2,dc=internal
  ```

The example configurations in Step 3: Create a configuration JSON with Presto properties for LDAP include this account for clarity, with the exception of the 5.10.0 example, where it is not supported. If the LDAP server uses anonymous binding, remove the `internal-communication.authentication.ldap.user` and `internal-communication.authentication.ldap.password` name/value pairs.

The LDAP distinguished name (DN) for Presto users

When you specify the LDAP configuration for Presto, you specify a bind pattern that consists of `${USER}` along with an organizational unit (OU) and additional domain components (DCs). Presto replaces `${USER}` with the actual User ID (UID) of each user during password authentication to match the distinguished name (DN) that this bind pattern specifies. You need the OUs that eligible users belong to and their DCs. For example, to allow users from the `admins` OU in the `corp.example.com` domain to authenticate to Presto, you specify `${USER},ou=admins,dc=corp,dc=example,dc=com` as the user bind pattern.

> **Note**
>
> When you use AWS CloudFormation, you need to use the Fn::Sub function in order to replace `${USER}` with the actual User ID (UID). For more information, see the Fn::Sub topic in the *AWS CloudFormation User Guide*.

When using Amazon EMR 5.10.0, you can specify only one such pattern. Using Amazon EMR 5.11.0 or later, you can specify multiple patterns separated by a colon (:). Users attempting to authenticate to Presto are compared to the first pattern, then the second, and so on. For an example, see Step 3: Create a configuration JSON with Presto properties for LDAP.

Step 2: Set up a security configuration

Create a security configuration with in-transit encryption enabled. For more information, see Create a security configuration in the *Amazon EMR Management Guide*. The encryption artifacts that you provide when you set up in-transit encryption are used to encrypt internal communication between Presto nodes. For more information, see Providing certificates for in-transit data encryption. The LDAP server certificate is used to authenticate client connections to the Presto server.

Step 3: Create a configuration JSON with Presto properties for LDAP

You use the `presto-config` configuration classification to set Presto properties for LDAP. The format and contents of `presto-config` are slightly different depending on the Amazon EMR release version and the Presto installation (PrestoDB or Trino). Examples of configuration differences are provided later in this section. For more information, see Configure applications.

The following steps assume that you save the JSON data to a file, *`MyPrestoConfig.json`*. If you use the console, upload the file to a secure location in Amazon S3 so that you can reference it when you create the cluster. If you use the AWS CLI, you can reference the file locally.

Example Amazon EMR 6.1.0 and later with PrestoSQL (Trino)

The following example uses the LDAP host name from Step 1: Gather information about your LDAP server and copy the server certificate to Amazon S3 to authenticate to the LDAP server for binding. Two user bind patterns are specified, which indicates that users within the `admins` OU and the `datascientists` OU on the LDAP server are eligible for authentication to the Trino server as users. The bind patterns are separated by a colon (`:`).

Amazon EMR versions 6.4.0 and later use the new name Trino instead of PrestoSQL. If you use Trino, replace *`prestosql-config`* in the following configuration classification with `trino-config` and *`prestosql-password-authenticator`* with `trino-password-authenticator`.

```
[
  {
```

```
      "Classification":"prestosql-config",
      "Properties":{
         "http-server.authentication.type":"PASSWORD"
      }
   },
   {
      "Classification":"prestosql-password-authenticator",
      "Properties":{
         "password-authenticator.name":"ldap",
         "ldap.url":"ldaps://ip-xxx-xxx-xxx-xxx.ec2.internal:636",
         "ldap.user-bind-pattern": "uid=${USER},ou=admins,dc=ec2,dc=internal:uid=
${USER},ou=datascientists,dc=ec2,dc=internal"
      }
   }
]
```

Example Amazon EMR 5.16.0 and later

The following example uses the LDAP user ID and password, and the LDAP host name from Step 1: Gather information about your LDAP server and copy the server certificate to Amazon S3 to authenticate to the LDAP server for binding. Two user bind patterns are specified, which indicates that users within the `admins` OU and the `datascientists` OU on the LDAP server are eligible for authentication to the Presto server as users. The bind patterns are separated by a colon (`:`).

```
[{
        "Classification": "presto-config",
                "Properties": {
                        "http-server.authentication.type": "PASSWORD"
                }
        },
        {
                "Classification": "presto-password-authenticator",
                "Properties": {
                        "password-authenticator.name": "ldap",
                        "ldap.url": "ldaps://ip-xxx-xxx-xxx-xxx.ec2.internal:636",
                        "ldap.user-bind-pattern": "uid=
${USER},ou=admins,dc=ec2,dc=internal:uid=${USER},ou=datascientists,dc=ec2,dc=internal",
                        "internal-communication.authentication.ldap.user": "presto",
                        "internal-communication.authentication.ldap.password": "123456"
                }
        }]
```

Example Amazon EMR 5.11.0 through 5.15.0

The format of the `presto-config` configuration classification is slightly different for these release versions. The following example specifies the same parameters as the previous example.

```
[{
        "Classification": "presto-config",
                "Properties": {
                        "http-server.authentication.type": "LDAP",
                        "authentication.ldap.url": "ldaps://ip-xxx-xxx-xxx-
xxx.ec2.internal:636",
                        "authentication.ldap.user-bind-pattern": "uid=
${USER},ou=admins,dc=ec2,dc=internal:uid=${USER},ou=datascientists,dc=ec2,dc=internal",
                        "internal-communication.authentication.ldap.user": "presto",
                        "internal-communication.authentication.ldap.password": "123456"
                }
        }]
```

Example Amazon EMR 5.10.0

Amazon EMR 5.10.0 supports anonymous binding only, so those entries are omitted. In addition, only a single bind pattern can be specified.

```
[{
        "Classification": "presto-config",
                "Properties": {
                        "http-server.authentication.type": "LDAP",
                        "authentication.ldap.url": "ldaps://ip-xxx-xxx-xxx-
xxx.ec2.internal:636",
                        "ldap.user-bind-pattern": "uid=
${USER},ou=prestousers,dc=ec2,dc=internal"
                }
        }]
```

Step 4: Create the script to copy the LDAP server certificate and upload it to Amazon S3

Create a script that copies the certificate file to each node in the cluster and adds it to the keystore. Create the script using a text editor, save it, and then upload it to Amazon S3. In Step 5: Create the cluster, the script file is referenced as *`s3://MyBucket/LoadLDAPCert.sh`*.

The following example script uses the default keystore password, *changeit*. We recommend that you connect to the master node after you create the cluster and change the keystore password using the keytool command.

```
#!/bin/bash
aws s3 cp s3://MyBucket/ldap_server.crt .
sudo keytool -import -keystore /usr/lib/jvm/jre-1.8.0-openjdk.x86_64/lib/security/
cacerts -trustcacerts -alias ldap_server -file ./ldap_server.crt -storepass changeit -
noprompt
```

Step 5: Create the cluster

When you create the cluster, you specify Presto and other applications that you want Amazon EMR to install. The following examples also reference the configuration classification properties within a JSON, but you can also specify the configuration classification inline.

To create a Presto cluster with LDAP authentication using the Amazon EMR console

1. Navigate to the new Amazon EMR console and select **Switch to the old console** from the side navigation. For more information on what to expect when you switch to the old console, see Using the old console.
2. Choose **Create cluster**, **Go to advanced options**.
3. Choose **Presto** along with other applications for Amazon EMR to install, and under **Software Configuration**, select the **Release** of Amazon EMR to use. LDAP authentication is supported only with Amazon EMR 5.10.0 and later.
4. Under **Edit software settings**, choose **Load JSON from S3** , enter the location in Amazon S3 of the JSON configuration file you created in Step 3: Create a configuration JSON with Presto properties for LDAP, and then choose **Next**.
5. Configure cluster hardware and networking, and then choose **Next**.
6. Choose **Bootstrap Actions**. For **Add bootstrap action**, select **Custom action**, and then choose **Configure and add**.
7. Enter a **Name** for the bootstrap action, enter the **Script location** that you created in Step 4: Create the script to copy the LDAP server certificate and upload it to Amazon S3, for example **s3://MyBucket/LoadLDAPCert.sh**, and then choose **Add**.
8. Under **General Options**, **Tags**, and **Additional Options** choose the settings that are appropriate for your application, and then choose **Next**.

9. Choose **Authentication and encryption**, and then select the **Security configuration** that you created in Step 2: Set up a security configuration.
10. Choose other security options as appropriate for your application, and then choose **Create cluster**.

To create a Presto cluster with LDAP authentication using the AWS CLI

- Use the `aws emr create-cluster` command. At a minimum, specify the Presto application, and also the Presto configuration classification, the bootstrap script, and the security configuration that you created in the previous steps. The following example references the configuration file as a JSON file saved in the same directory where you run the command. The bootstrap script, on the other hand, must be saved in Amazon S3. The following example uses `s3://MyBucket/LoadLDAPCert.sh`.

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster  --applications Name=presto --release-label emr-5.16.0 \
--use-default-roles --ec2-attributes KeyName=MyKeyPair,SubnetId=subnet-1234ab5
 \ --instance-count 3 --instance-type m5.xlarge --region us-west-2 --name
 "MyPrestoWithLDAPAuth" \
--bootstrap-actions Name="Distribute LDAP server cert",Path="s3://MyBucket/
LoadLDAPCert.sh" \
--security-configuration MyPrestoLDAPSecCfg --configurations file://
MyPrestoConfig.json
```

Activating Presto strict mode

In certain situations, long-running queries can lead to high costs and cause Amazon EMR to use more cluster resources. This takes resources away from other workloads on the cluster. With Amazon EMR versions 6.8 and later, you can use a strict mode feature that rejects or warns you about the following types of long-running queries:

- Queries without predicates on the partitioned columns that result in table scans of large amounts of data
- Queries with cross joins between two large tables
- Queries that sort large number of rows without limit

After Presto completely optimizes the query plan, strict mode runs. To use and customize strict mode to your query needs, you can configure Presto in the following ways.

Presto configurations for strict mode

Setting	Description	Default
`strict-mode-enabled`	Turns strict mode on and off. A value of `true` indicates that strict mode is on.	`false`
`strict-mode-fail-query`	Rejects queries if strict mode detects probable long-running queries. If `false`, Amazon EMR only raises a warning.	`false`
`strict-mode-restrictions`	Specifies the restrictions to apply when strict mode is turned on. Strict mode supports the following restrictions: MANDATORY_PARTITION_PREDICATE, DISALLOW_CROSS_JOIN, and LIMITED_SORT.	MANDATORY_PARTITION_PREDICATE,DISALLOW_CROSS_JOIN, LIMITED_SORT

To experiment with strict mode, you can override these configurations, or set them as session properties when you use the Presto client.

To set the configuration at cluster creation with the AWS Management Console

1. Choose **Create cluster** and select Amazon EMR version 6.8.0, and Presto or Trino. For more information, see Installing PrestoDB and Trino.

2. Specify the configuration properties for strict mode directly, or upload a JSON file to Amazon S3. Optionally, select the for your metastore. Specify your VPC, subnets, bootstrap actions, key pair, and security group. Choose **Create cluster** to create your cluster.
3. Log in to the primary node of the cluster and run `presto-cli` or `trino-cli`.
4. Submit your queries. Strict mode validates each query and determines if it is long-running. Depending on your `strict-mode-fail-query` setting, Amazon EMR rejects the query or raises a warning.
5. When you're finished with your queries, terminate the cluster and delete your resources.

To set the configuration on a running cluster with the AWS CLI

1. Log in to the primary node of your cluster with the AWS CLI and run `presto-cli` or `trino-cli`.
2. Run the following commands with your desired values.

```
set session strict_mode_enabled = true;
set session strict_mode_fail_query = false;
set session strict_mode_restrictions = 'DISALLOW_CROSS_JOIN,LIMITED_SORT';
```

Considerations

When you use strict mode, consider the following:

- In some cases, strict mode can reject short-running queries that don’t consume a lot of resources. For example, queries on small tables don’t apply dynamic filtering or replace inner joins with cross joins. This can lead the query to use the mandatory partition predicate or disallow cross join. When this happens, strict mode rejects the query.
- The strict mode check is only applied on SELECT, INSERT, CREATE TABLE AS SELECT, and EXPLAIN ANALYZE query types.
- You can only use the restriction on the mandatory partition predicate with the Hive connector.

Handling Spot Instance loss in Presto

With Spot Instances in Amazon EMR, you can run big data workloads on spare Amazon EC2 capacity at a reduced cost. In exchange for the lower cost, Amazon EC2 can interrupt Spot

Instances with a two-minute notification. When you terminate a node, Presto can take up to 10 minutes before it returns an error. This causes unnecessary delays in error reports and possible retries. Quick termination is a feature that gives you control over the way Presto handles terminated nodes.

The job of the Presto coordinator is to keep track of all the worker nodes with regular polls of their statuses. Without quick termination, the coordinator doesn't consult the YARN NodeManager for the status of each node. This can result in a long retry loop before the query fails. With quick termination, the Presto coordinator consults the node status in the NodeManager as soon as the poll fails to reach the host. If NodeManager shows that the node is inactive, Presto abandons further retries, fails the query, and returns a `NODE_DECOMMISSIONED` error.

The following set of configuration parameters allows you to control and customize Presto behavior in the event of node termination.

Presto configurations for node failure handling

Setting	Description	Default
`query.remote-task.max-backoff-duration`	The duration of time that the coordinator continues attempts to fetch the remote task status from worker nodes.	10 minutes
`query.remote-task.quick-terminate-node-failure`	Activates quick node failure if the coordinator can't reach the node or can't connect to the worker that runs on that node. The value of `query.remote-task.terminate-on-connect-exception` determines if the coordinator must reach the node or connect to the worker. The node fails the query and Amazon EMR removes the	`true`

Setting	Description	Default
	node from the list of available workers. When this happens, you can't use the node to schedule new queries. When you set this value to `false`, Presto reverts to its previous behavior where the Presto coordinator again tries to reach the node (for `query.remote-task.max-backoff-duration`) before it marks the node as unavailable and fails the ongoing query on the node.	
`query.remote-task.terminate-on-connect-exception`	Specifies if Amazon EMR should a node if the host is reachable but the coordinator fails to connect to the worker process of the host. When you set this value to `true`, you activate quick query failure if the host is unreachable.	`false`

Fault-tolerant execution in Trino

Fault-tolerant execution is a mechanism in Trino that a cluster can use to mitigate query failures. To do this, it retries queries or their component tasks when they fail. When fault-tolerant execution is activated, intermediate exchange data is spooled, and another worker can reuse it in the event of a worker outage or other fault during query execution.

For more information on fault-tolerant execution in Trino, see Project Tardigrade delivers ETL at Trino speeds to early users on the Trino blog.

Configuration

Fault-tolerant execution is deactivated by default. To activate the feature, set the `retry-policy` configuration property in the `trino-config` classification to either QUERY or TASK based on the desired retry policy, as follows.

```
{"classification":
    "trino-config",
"properties":
    {
    "retry-policy":
        "QUERY"
    }
}
```

A **QUERY** retry policy instructs Trino to retry a query automatically when an error occurs on a worker node. We recommend that you use a QUERY retry policy when the majority of the workload for the Trino cluster comprises many small queries.

A **TASK** retry policy instructs Trino to retry individual query tasks in the event of failure. We recommend this policy when Trino executes large batch queries. The cluster can more efficiently retry smaller tasks within the query rather than retry the whole query.

Exchange manager

An exchange manager stores and manages spooled data for fault-tolerant execution. It uses external storage to store spilled data beyond the in-memory buffer size. You can configure a file system-based exchange manager that stores spooled data in a specified location, such as Amazon S3, Amazon S3 compatible systems, or HDFS.

Amazon EMR releases 6.9.0 and later include the `trino-exchange-manager` classification to configure the exchange manager. These releases also support HDFS for spooling.

Setting up exchange manager

Use the `trino-exchange-manager` configuration classification to configure an exchange manager. This classification internally creates a `etc/exchange-manager.properties` configuration file on the coordinator and all worker nodes. The classification also sets `exchange-manager.name` configuration property to `filesystem`.

By default, Amazon EMR releases 6.9.0 and later use HDFS as an exchange manager. HDFS is available in the Amazon EMR EC2 clusters, and spooling occurs in the `trino-exchange/` directory by default. To use the default settings, set the following configuration:

```
{"Classification":
    "trino-exchange-manager"
}
```

If you want to provide a custom location, set the following properties in the `trino-exchange-manager` classification:

- Set `exchange.use-local-hdfs` to `true`.
- Set `exchange.base-directories` to the custom directory location in HDFS, for example, `exchange.base-directories=/exchange`. If the custom directory isn't already in HDFS, Amazon EMR will create it.

HDFS exchange manager configurations

Based on internal testing results, we recommend that you spool to local HDFS for better query performance in comparison to other cloud-based file systems. You can set the following configurations for the exchange manager with HDFS.

Configuration	Description	Default setting
`exchange.hdfs.block-size`	Block size for HDFS storage	4 MB
`hdfs.config.resources`	List of file paths to configure HDFS	If `exchange.use-local-hdfs` is `true`, uses the paths to `core-site.xml` , `hdfs-site.xml` files; otherwise `null`

For additional fault-tolerant execution configuration properties, and for information on how to set up Amazon S3 or other Amazon S3 compatible systems for spooling, see the [Fault-tolerant execution](#) page of the Trino documentation.

Considerations and limitations

- If you enable fault-tolerant execution, it disables `write` operations for connectors that do not support `write` when `retry-policy` is set. As of Amazon EMR release 6.9.0, Delta Lake, Hive and Iceberg connectors support `write` operations with `retry-policy`.
- If you use exchange manager and perform expensive I/O operations, your queries may experience degraded performance while exchange manager spools the intermediate data to external storage.

Using Presto automatic scaling with Graceful Decommission

Amazon EMR release versions 5.30.0 and later include a feature you can use to set a grace period for certain scaling actions. The grace period allows Presto tasks to keep running before the node terminates because of a scale-in resize action or an automatic scaling policy request. For more information about scaling rules, see Understanding automatic scaling rules in the *Amazon EMR Management Guide*. Presto autoscaling with Graceful Decommission prevents new tasks from being scheduled on a node that is decommissioning, while at the same time allowing tasks that are already running to complete before the shut down timeout is reached. Queries that are running will complete execution before the node is decommissioned. Autoscaling is not supported on instance fleets.

You can control how much time to allow for Presto tasks to complete after an autoscale shut down request is received. By default, the shut down timeout for Amazon EMR is `0` minutes, which means that Amazon EMR immediately terminates the node and any Presto tasks running on it, if required by a scale-in request. To set a longer timeout for Presto tasks on Amazon EMR to allow running queries to complete before scaling down a cluster, use the `presto-config` configuration classification to set the `graceful-shutdown-timeout` parameter to a value in seconds or minutes greater than zero. For more information, see Configure applications.

For example, increasing the `graceful-shutdown-timeout` value to `"30m"` specifies a timeout period of 30 minutes. After the shut down timeout period ends, the node marked for decommissioning is forcefully terminated if it is waiting for query tasks to complete, and the query fails. If the query tasks finish in five minutes, the node marked for decommissioning terminates at five minutes, provided other YARN applications have completed execution.

Example Example Presto autoscale configuration with Graceful Decommission

Replace the `graceful-shutdown-timeout` value with the number of minutes appropriate for your setup. There's no maximum value. The example below sets a timeout value of `1800` seconds (30 minutes).

```
[
    {
        "classification": "presto-config",
        "properties": {
            "graceful-shutdown-timeout": "1800s"
        }
    }
]
```

Limitations

PrestoDB Graceful Decommission does not work on EMR clusters where HTTP connectivity is disabled, such as when `http-server.http.enabled` is set to `false`. Trino does not support Graceful Decommission at all, regardless of the `http-server.http.enabled` setting.

Considerations with Presto on Amazon EMR

Consider the following limitations when you run Presto on Amazon EMR.

Presto command line executable

In Amazon EMR, PrestoDB and Trino both use the same command line executable, `presto-cli`, as in the following example.

```
presto-cli --catalog hive
```

Non-configurable Presto deployment properties

The version of Amazon EMR that you use determines the Presto deployment configurations that are available. For more information about these configuration properties, see Deploying Presto in the Presto documentation. The following table shows the different configuration options for Presto `properties` files.

File	Configurable
`log.properties`	PrestoDB: Configurable in Amazon EMR versions 4.0.0 and later. Use the `presto-log` configuration classification. Trino (PrestoSQL): Configurable in Amazon EMR versions 6.1.0 and later. Use the `prestosql-log` or `trino-log` configuration classification.
`config.properties`	PrestoDB: Configurable in Amazon EMR versions 4.0.0 and later. Use the `presto-config` configuration classification. Trino (PrestoSQL): Configurable in Amazon EMR versions 6.1.0 and later. Use the `prestosql-config` or `trino-config` configuration classification.
`hive.properties`	PrestoDB: Configurable in Amazon EMR versions 4.1.0 and later. Use the `presto-connector-hive` configuration classification. Trino (PrestoSQL): Configurable in Amazon EMR versions 6.1.0 and later. Use the `prestosql-connector-hive` or `trino-connector-hive` configuration classification.
`node.properties`	PrestoDB: Configurable in Amazon EMR version 5.6.0 and later. Use the `presto-node` configuration classification. Trino (PrestoSQL): Configurable in Amazon EMR versions 6.1.0 and later. Use the `prestosql-node` or `trino-node` configuration classification.

File	Configurable
`jvm.config`	Not configurable.

PrestoDB and Trino installation

The application name *Presto* continues to be used to install PrestoDB on clusters. To install Trino on clusters, use the application name *Trino* (or *PrestoSQL* in older versions of Amazon EMR).

You can install either PrestoDB or Trino, but you can't install both on a single cluster. If you specify both PrestoDB and Trino when you attempt to create a cluster, a validation error occurs and the cluster creation request fails.

EMRFS and PrestoS3FileSystem configuration

With Amazon EMR versions 5.12.0 and later, PrestoDB can use EMRFS. This is the default configuration. EMRFS is also the default file system for Trino (PrestoSQL) in Amazon EMR versions 6.1.0 and later. For more information, see EMR File System (EMRFS) in the *Amazon EMR Management Guide*. With earlier versions of Amazon EMR, PrestoS3FileSystem is the only configuration option.

You can use a security configuration to set up encryption for EMRFS data in Amazon S3. You can also use IAM roles for EMRFS requests to Amazon S3. For more information, see Understanding encryption options and Configure IAM roles for EMRFS requests to Amazon S3 in the *Amazon EMR Management Guide*.

> **Note**
>
> If you query underlying data in Amazon S3 with Amazon EMR version 5.12.0, Presto errors can occur. This is because Presto fails to pick up configuration classification values from `emrfs-site.xml`. As a workaround, create an `emrfs` subdirectory under `usr/lib/presto/plugin/hive-hadoop2/` and create a symlink in `usr/lib/presto/plugin/hive-hadoop2/emrfs` to the existing `/usr/share/aws/emr/emrfs/conf/emrfs-site.xml` file. Then restart the presto-server process (`sudo presto-server stop` followed by `sudo presto-server start`).

You can override the EMRFS default and use the PrestoS3FileSystem instead. To do this, use the `presto-connector-hive` configuration classification to set `hive.s3-file-system-type` to PRESTO as shown in the following example. For more information, see [Configure applications](#).

```
[
   {
      "Classification": "presto-connector-hive",
      "Properties": {
         "hive.s3-file-system-type": "PRESTO"
      }
   }
]
```

If you use PrestoS3FileSystem, use the `presto-connector-hive` configuration classification or `trino-connector-hive` for Trino to configure PrestoS3FileSystem properties. For more information about available properties, see [Amazon S3 configuration](#) in the Hive Connector section of the Presto documentation. These settings do not apply to EMRFS.

Default setting for end user impersonation

By default, Amazon EMR versions 5.12.0 and later enable end user impersonation for access to HDFS. For more information, see [End user impersonation](#) in the Presto documentation. To change this setting with the `presto-config` configuration classification, set the `hive.hdfs.impersonation.enabled` property to `false`.

Default port for Presto web interface

By default, Amazon EMR configures the Presto web interface on the Presto coordinator to use port 8889 (for PrestoDB and Trino). To change the port, use the `presto-config` configuration classification to set the `http-server.http.port` property. For more information, see [Config properties](#) in the *Deploying Presto* section of Presto Documentation.

Issue with Hive Bucket execution in some releases

Presto version 152.3 has an issue with Hive bucket execution that causes significantly slower Presto query performance in some circumstances. Amazon EMR versions 5.0.3, 5.1.0, and 5.2.0 include this version of Presto. To mitigate this issue, use the `presto-connector-hive` configuration classification to set the `hive.bucket-execution` property to `false`, as shown in the following example.

```
[
   {
      "Classification": "presto-connector-hive",
      "Properties": {
         "hive.bucket-execution": "false"
      }
   }
]
```

Presto release history

The following table lists the version of Presto included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in Amazon EMR 7.x release versions, Amazon EMR 6.x release versions, or Amazon EMR 5.x release versions.

Presto version information

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-7.2.0	0.285	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.36.2	0.267	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-

Amazon EMR Release label	Presto Version	Components installed with Presto
		yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server , hive-client, hudi, hudi-pres to, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-7.1.0	0.284	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server , hive-client, hudi, hudi-pres to, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-7.0.0	0.283	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server , hive-client, hudi, hudi-pres to, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-6.15.0	0.283	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.14.0	0.281	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-6.13.0	0.281	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.12.0	0.281	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-6.11.1	0.279	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.11.0	0.279	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-6.10.1	0.278	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.10.0	0.278	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-6.9.1	0.276	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.9.0	0.276	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-6.8.1	0.273	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.8.0	0.273	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-6.7.0	0.272	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.36.1	0.267	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.36.0	0.267	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.6.0	0.267	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.35.0	0.266	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.5.0	0.261	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-6.4.0	0.254.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.3.1	0.245.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-6.3.0	0.245.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.2.1	0.238.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-6.2.0	0.238.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.1.1	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-6.1.0	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.0.1	0.230	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-6.0.0	0.230	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.34.0	0.261	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.33.1	0.245.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.33.0	0.245.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.32.1	0.240.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.32.0	0.240.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.31.1	0.238.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.31.0	0.238.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.30.2	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.30.1	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.30.0	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.29.0	0.227	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.28.1	0.227	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.28.0	0.227	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.27.1	0.224	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.27.0	0.224	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.26.0	0.220	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.25.0	0.220	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.24.1	0.219	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.24.0	0.219	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.23.1	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.23.0	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.22.0	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.21.2	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.21.1	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.21.0	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.20.1	0.214	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.20.0	0.214	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.19.1	0.212	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.19.0	0.212	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.18.1	0.210	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.18.0	0.210	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.17.2	0.206	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.17.1	0.206	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.17.0	0.206	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.16.1	0.203	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.16.0	0.203	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.15.1	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.15.0	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.14.2	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.14.1	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.14.0	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.13.1	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.13.0	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.12.3	0.188	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.12.2	0.188	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.12.1	0.188	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.12.0	0.188	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.11.4	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.11.3	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.11.2	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.11.1	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.11.0	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.10.1	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.10.0	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.9.1	0.184	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.9.0	0.184	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.8.3	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.8.2	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.8.1	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.8.0	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.7.1	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.7.0	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.6.1	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.6.0	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.5.4	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.5.3	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.5.2	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.5.1	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.5.0	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.4.1	0.166	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.4.0	0.166	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.3.2	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.3.1	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.3.0	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.2.3	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.2.2	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.2.1	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.2.0	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.1.1	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.1.0	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.0.3	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release label	Presto Version	Components installed with Presto
emr-5.0.2	0.150	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.0.1	0.150	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.0.0	0.150	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

The following table lists the version of Trino (Presto SQL) included in each release version of Amazon EMR, along with the components installed with the application. PrestoSQL changed its name to Trino starting with version 351.

Trino (PrestoSQL) version information

Amazon EMR Release label	Trino (PrestoSQL) Version	Components installed with Trino (PrestoSQL)
emr-7.2.0	436	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-7.1.0	435	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-7.0.0	426	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library,

Amazon EMR Release label	Trino (PrestoSQL) Version	Components installed with Trino (PrestoSQL)
		hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server , hive-client, hudi, hudi-trin o, hcatalog-server, mariadb-s erver, trino-coordinator, trino-worker
emr-6.15.0	426	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server , hive-client, hudi, hudi-trin o, hcatalog-server, mariadb-s erver, trino-coordinator, trino-worker

Amazon EMR Release label	Trino (PrestoSQL) Version	Components installed with Trino (PrestoSQL)
emr-6.14.0	422	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.13.0	414	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Amazon EMR Release label	Trino (PrestoSQL) Version	Components installed with Trino (PrestoSQL)
emr-6.12.0	414	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.11.1	410	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Amazon EMR Release label	Trino (PrestoSQL) Version	Components installed with Trino (PrestoSQL)
emr-6.11.0	410	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.10.1	403	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Amazon EMR Release label	Trino (PrestoSQL) Version	Components installed with Trino (PrestoSQL)
emr-6.10.0	403	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.9.1	398	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Amazon EMR Release label	Trino (PrestoSQL) Version	Components installed with Trino (PrestoSQL)
emr-6.9.0	398	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.8.1	388	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Amazon EMR Release label	Trino (PrestoSQL) Version	Components installed with Trino (PrestoSQL)
emr-6.8.0	388	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.7.0	378	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Amazon EMR Release label	Trino (PrestoSQL) Version	Components installed with Trino (PrestoSQL)
emr-6.6.0	367	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.5.0	360	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Amazon EMR Release label	Trino (PrestoSQL) Version	Components installed with Trino (PrestoSQL)
emr-6.4.0	359	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.3.1	350	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker

Amazon EMR Release label	Trino (PrestoSQL) Version	Components installed with Trino (PrestoSQL)
emr-6.3.0	350	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker
emr-6.2.1	343	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker

Amazon EMR Release label	Trino (PrestoSQL) Version	Components installed with Trino (PrestoSQL)
emr-6.2.0	343	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker
emr-6.1.1	338	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker

Amazon EMR Release label	Trino (PrestoSQL) Version	Components installed with Trino (PrestoSQL)
emr-6.1.0	338	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker

Trino (PrestoSQL) release notes by version

- Amazon EMR 6.9.0 - Trino (PrestoSQL) release notes

Amazon EMR 6.9.0 - Trino (PrestoSQL) release notes

Amazon EMR 6.9.0 - Trino (PrestoSQL) new features

- To support long running queries, Trino now includes a fault-tolerant execution mechanism. Fault-tolerant execution mitigates query failures by retrying failed queries or their component tasks. For more information, see Fault-tolerant execution in Trino.

Amazon EMR 6.9.0 - Trino (PrestoSQL) changes

Amazon EMR 6.9.0 - PrestoDB changes

Type	Description
Upgrade	PrestoDB upgrade to 0.276
Upgrade	Support for Hadoop 3.3.3

Type	Description
Upgrade	Hudi upgrade to 0.12.1
Feature	Amazon EMR and Presto integration with AWS Lake Formation for interactive workloads using GCSC API.
Feature	Added Kerberos-related configs inside Security Configs for PrestoDB to enable Keberos.
Bug fix	Reverted OSS pull request #18115 which was added to reduce the number of `hdfsConfiguration` copies. This caused the HDFS configuration copy incorrectly when using EMRFS or Hudi tables.

Amazon EMR 6.9.0 - Trino changes

Type	Description
Upgrade	Trino Upgrade to 398
Upgrade	Support for Hadoop 3.3.3
Feature	Tardigrade support: Add support for exchange spooling on HDFS and Amazon S3. For more information, see Fault-tolerant execution in Trino
Bug fix	When Trino Iceberg is used and Glue catalog is enabled, avoid adding metastore uri in `iceberg.properties.`

Amazon EMR 6.9.0 - Trino (PrestoSQL) known issues

- For Amazon EMR release 6.9.0, Trino does not work on clusters enabled for Apache Ranger. If you need to use Trino with Ranger, contact AWS Support.

Apache Spark

Apache Spark is a distributed processing framework and programming model that helps you do machine learning, stream processing, or graph analytics with Amazon EMR clusters. Similar to Apache Hadoop, Spark is an open-source, distributed processing system commonly used for big data workloads. However, Spark has several notable differences from Hadoop MapReduce. Spark has an optimized directed acyclic graph (DAG) execution engine and actively caches data in-memory, which can boost performance, especially for certain algorithms and interactive queries.

Spark natively supports applications written in Scala, Python, and Java. It also includes several tightly integrated libraries for SQL (Spark SQL), machine learning (MLlib), stream processing (Spark streaming), and graph processing (GraphX). These tools make it easier to leverage the Spark framework for a wide variety of use cases.

You can install Spark on an Amazon EMR cluster along with other Hadoop applications, and it can also leverage the Amazon EMR file system (EMRFS) to directly access data in Amazon S3. Hive is also integrated with Spark so that you can use a HiveContext object to run Hive scripts using Spark. A Hive context is included in the spark-shell as `sqlContext`.

For an example tutorial on setting up an EMR cluster with Spark and analyzing a sample data set, see Tutorial: Getting started with Amazon EMR on the AWS News blog.

Important

Apache Spark version 2.3.1, available beginning with Amazon EMR release 5.16.0, addresses CVE-2018-8024 and CVE-2018-1334. We recommend that you migrate earlier versions of Spark to Spark version 2.3.1 or later.

The following table lists the version of Spark included in the latest release of the Amazon EMR 7.x series, along with the components that Amazon EMR installs with Spark.

For the version of components installed with Spark in this release, see Release 7.2.0 Component Versions.

Spark version information for emr-7.2.0

Amazon EMR Release Label	Spark Version	Components Installed With Spark
emr-7.2.0	Spark 3.5.1	delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

The following table lists the version of Spark included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with Spark.

For the version of components installed with Spark in this release, see Release 6.15.0 Component Versions.

Spark version information for emr-6.15.0

Amazon EMR Release Label	Spark Version	Components Installed With Spark
emr-6.15.0	Spark 3.4.1	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server,

Amazon EMR Release Label	Spark Version	Components Installed With Spark
		hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Note

Amazon EMR release 6.8.0 comes with Apache Spark 3.3.0. This Spark release uses Apache Log4j 2 and the `log4j2.properties` file to configure Log4j in Spark processes. If you use Spark in the cluster or create EMR clusters with custom configuration parameters, and you want to upgrade to Amazon EMR release 6.8.0, you must migrate to the new `spark-log4j2` configuration classification and key format for Apache Log4j 2. For more information, see Migrating from Apache Log4j 1.x to Log4j 2.x.

The following table lists the version of Spark included in the latest release of the Amazon EMR 5.x series, along with the components that Amazon EMR installs with Spark.

For the version of components installed with Spark in this release, see Release 5.36.2 Component Versions.

Spark version information for emr-5.36.2

Amazon EMR Release Label	Spark Version	Components Installed With Spark
emr-5.36.2	Spark 2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-

Amazon EMR Release Label	Spark Version	Components Installed With Spark
		hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Topics

- Create a cluster with Apache Spark
- Run Spark applications with Docker on Amazon EMR 6.x
- Use the AWS Glue Data Catalog as the metastore for Spark SQL
- Configure Spark
- Optimize Spark performance
- Spark Result Fragment Caching
- Use the Nvidia RAPIDS Accelerator for Apache Spark
- Access the Spark shell
- Use Amazon SageMaker Spark for machine learning
- Write a Spark application
- Improve Spark performance with Amazon S3
- Add a Spark step
- View Spark application history
- Access the Spark web UIs
- Using the Spark structured streaming Amazon Kinesis Data Streams connector
- Using Amazon Redshift integration for Apache Spark with Amazon EMR
- Spark release history

Create a cluster with Apache Spark

The following procedure creates a cluster with Spark installed using **Quick Options** in the Amazon EMR console.

You can alternatively use **Advanced Options** to further customize your cluster setup, or to submit steps to programmatically install applications and then run custom applications. With either cluster creation option, you can choose to use AWS Glue as your Spark SQL metastore. See Use the AWS Glue Data Catalog as the metastore for Spark SQL for more information.

To launch a cluster with Spark installed

1. Open the Amazon EMR console at https://console.aws.amazon.com/emr.
2. Choose **Create cluster** to use **Quick Options**.
3. Enter a **Cluster name**. Your cluster name can't contain the characters <, >, $, |, or ` (backtick).
4. For **Software Configuration**, choose a **Release** option.
5. For **Applications**, choose the **Spark** application bundle.
6. Select other options as necessary and then choose **Create cluster**.

> **Note**
>
> To configure Spark when you are creating the cluster, see Configure Spark.

To launch a cluster with Spark installed using the AWS CLI

- Create the cluster with the following command.

```
aws emr create-cluster --name "Spark cluster" --release-label emr-7.2.0 --
applications Name=Spark \
--ec2-attributes KeyName=myKey --instance-type m5.xlarge --instance-count 3 --use-
default-roles
```

Note

Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

To launch a cluster with Spark installed using the SDK for Java

Specify Spark as an application with `SupportedProductConfig` used in `RunJobFlowRequest`.

- The following example shows how to create a cluster with Spark using Java.

```
import com.amazonaws.AmazonClientException;
import com.amazonaws.auth.AWSCredentials;
import com.amazonaws.auth.AWSStaticCredentialsProvider;
import com.amazonaws.auth.profile.ProfileCredentialsProvider;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduce;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduceClientBuilder;
import com.amazonaws.services.elasticmapreduce.model.*;
import com.amazonaws.services.elasticmapreduce.util.StepFactory;

public class Main {

        public static void main(String[] args) {
                AWSCredentials credentials_profile = null;
                try {
                        credentials_profile = new
 ProfileCredentialsProvider("default").getCredentials();
                } catch (Exception e) {
                        throw new AmazonClientException(
                                        "Cannot load credentials from .aws/
credentials file. " +
                                                        "Make sure that the
 credentials file exists and the profile name is specified within it.",
                                        e);
                }

                AmazonElasticMapReduce emr =
 AmazonElasticMapReduceClientBuilder.standard()
                                .withCredentials(new
 AWSStaticCredentialsProvider(credentials_profile))
                                .withRegion(Regions.US_WEST_1)
```

```
                                .build();

                // create a step to enable debugging in the AWS Management Console
                StepFactory stepFactory = new StepFactory();
                StepConfig enabledebugging = new StepConfig()
                                .withName("Enable debugging")
                                .withActionOnFailure("TERMINATE_JOB_FLOW")

 .withHadoopJarStep(stepFactory.newEnableDebuggingStep());

                Application spark = new Application().withName("Spark");

                RunJobFlowRequest request = new RunJobFlowRequest()
                                .withName("Spark Cluster")
                                .withReleaseLabel("emr-5.20.0")
                                .withSteps(enabledebugging)
                                .withApplications(spark)
                                .withLogUri("s3://path/to/my/logs/")
                                .withServiceRole("EMR_DefaultRole")
                                .withJobFlowRole("EMR_EC2_DefaultRole")
                                .withInstances(new JobFlowInstancesConfig()
                                                .withEc2SubnetId("subnet-12ab3c45")
                                                .withEc2KeyName("myEc2Key")
                                                .withInstanceCount(3)

 .withKeepJobFlowAliveWhenNoSteps(true)
                                                .withMasterInstanceType("m4.large")

 .withSlaveInstanceType("m4.large"));
                RunJobFlowResult result = emr.runJobFlow(request);
                System.out.println("The cluster ID is " + result.toString());
        }
}
```

Run Spark applications with Docker on Amazon EMR 6.x

With Amazon EMR 6.0.0, Spark applications can use Docker containers to define their library dependencies, instead of installing dependencies on the individual Amazon EC2 instances in the cluster. To run Spark with Docker, you must first configure the Docker registry and define additional parameters when submitting a Spark application. For more information, see Configure Docker integration.

When the application is submitted, YARN invokes Docker to pull the specified Docker image and run the Spark application inside a Docker container. This allows you to easily define and isolate dependencies. It reduces the time for bootstrapping or preparing instances in the Amazon EMR cluster with the libraries needed for job execution.

Considerations when running Spark with Docker

When you run Spark with Docker, make sure the following prerequisites are met:

- The `docker` package and CLI are only installed on core and task nodes.
- On Amazon EMR 6.1.0 and later, you can alternatively install Docker on a primary node by using following commands.
  - ```
    sudo yum install -y docker
    sudo systemctl start docker
    ```
- The `spark-submit` command should always be run from a primary instance on the Amazon EMR cluster.
- The Docker registries used to resolve Docker images must be defined using the Classification API with the `container-executor` classification key to define additional parameters when launching the cluster:
  - `docker.trusted.registries`
  - `docker.privileged-containers.registries`
- To execute a Spark application in a Docker container, the following configuration options are necessary:
  - `YARN_CONTAINER_RUNTIME_TYPE=docker`
  - `YARN_CONTAINER_RUNTIME_DOCKER_IMAGE={DOCKER_IMAGE_NAME}`
- When using Amazon ECR to retrieve Docker images, you must configure the cluster to authenticate itself. To do so, you must use the following configuration option:
  - `YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG={DOCKER_CLIENT_CONFIG_PATH_ON_HDFS`
- In Amazon EMR 6.1.0 and later, you are not required to use the listed command `YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG={DOCKER_CLIENT_CONFIG_PATH_ON_HDFS` when the ECR auto authentication feature is enabled.
- Any Docker image used with Spark must have Java installed in the Docker image.

For more information about the prerequisites, see Configure Docker integration.
    ```

Creating a Docker image

Docker images are created using a Dockerfile, which defines the packages and configuration to include in the image. The following two example Dockerfiles use PySpark and SparkR.

PySpark Dockerfile

Docker images created from this Dockerfile include Python 3 and the NumPy Python package. This Dockerfile uses Amazon Linux 2 and the Amazon Corretto JDK 8.

```
FROM amazoncorretto:8

RUN yum -y update
RUN yum -y install yum-utils
RUN yum -y groupinstall development

RUN yum list python3*
RUN yum -y install python3 python3-dev python3-pip python3-virtualenv

RUN python -V
RUN python3 -V

ENV PYSPARK_DRIVER_PYTHON python3
ENV PYSPARK_PYTHON python3

RUN pip3 install --upgrade pip
RUN pip3 install numpy pandas

RUN python3 -c "import numpy as np"
```

SparkR Dockerfile

Docker images created from this Dockerfile include R and the randomForest CRAN package. This Dockerfile includes Amazon Linux 2 and the Amazon Corretto JDK 8.

```
FROM amazoncorretto:8

RUN java -version

RUN yum -y update
RUN amazon-linux-extras install R4
```

```
RUN yum -y install curl hostname

#setup R configs
RUN echo "r <- getOption('repos'); r['CRAN'] <- 'http://cran.us.r-project.org';
 options(repos = r);" > ~/.Rprofile

RUN Rscript -e "install.packages('randomForest')"
```

For more information on Dockerfile syntax, see the Dockerfile reference documentation.

Using Docker images from Amazon ECR

Amazon Elastic Container Registry (Amazon ECR) is a fully-managed Docker container registry, which makes it easy to store, manage, and deploy Docker container images. When using Amazon ECR, the cluster must be configured to trust your instance of ECR, and you must configure authentication in order for the cluster to use Docker images from Amazon ECR. For more information, see Configuring YARN to access Amazon ECR.

To make sure that Amazon EMR hosts can access the images stored in Amazon ECR, your cluster must have the permissions from the `AmazonEC2ContainerRegistryReadOnly` policy associated with the instance profile. For more information, see `AmazonEC2ContainerRegistryReadOnly` Policy.

In this example, the cluster must be created with the following additional configuration to ensure that the Amazon ECR registry is trusted. Replace the *`123456789123.dkr.ecr.us-east-1.amazonaws.com`* endpoint with your Amazon ECR endpoint.

```
[
  {
    "Classification": "container-executor",
    "Configurations": [
      {
        "Classification": "docker",
        "Properties": {
          "docker.privileged-containers.registries":
 "local,centos,123456789123.dkr.ecr.us-east-1.amazonaws.com",
          "docker.trusted.registries": "local,centos,123456789123.dkr.ecr.us-
east-1.amazonaws.com"
        }
      }
    ],
```

```
    "Properties": {}
  }
]
```

Using PySpark with Amazon ECR

The following example uses the PySpark Dockerfile, which will be tagged and uploaded to Amazon ECR. After you upload the Dockerfile, you can run the PySpark job and refer to the Docker image from Amazon ECR.

After you launch the cluster, use SSH to connect to a core node and run the following commands to build the local Docker image from the PySpark Dockerfile example.

First, create a directory and a Dockerfile.

```
mkdir pyspark
vi pyspark/Dockerfile
```

Paste the contents of the PySpark Dockerfile and run the following commands to build a Docker image.

```
sudo docker build -t local/pyspark-example pyspark/
```

Create the `emr-docker-examples` ECR repository for the examples.

```
aws ecr create-repository --repository-name emr-docker-examples
```

Tag and upload the locally built image to ECR, replacing *`123456789123.dkr.ecr.us-east-1.amazonaws.com`* with your ECR endpoint.

```
sudo docker tag local/pyspark-example 123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-examples:pyspark-example
sudo docker push 123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-examples:pyspark-example
```

Use SSH to connect to the primary node and prepare a Python script with the filename `main.py`. Paste the following content into the `main.py` file and save it.

```
from pyspark.sql import SparkSession
spark = SparkSession.builder.appName("docker-numpy").getOrCreate()
```

```
sc = spark.sparkContext

import numpy as np
a = np.arange(15).reshape(3, 5)
print(a)
```

On Amazon EMR 6.0.0, to submit the job, reference the name of the Docker image. Define the additional configuration parameters to make sure that the job execution uses Docker as the runtime. When using Amazon ECR, the `YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG` must reference the `config.json` file containing the credentials used to authenticate to Amazon ECR.

```
DOCKER_IMAGE_NAME=123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-
examples:pyspark-example
DOCKER_CLIENT_CONFIG=hdfs:///user/hadoop/config.json
spark-submit --master yarn \
--deploy-mode cluster \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG=
$DOCKER_CLIENT_CONFIG \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG=
$DOCKER_CLIENT_CONFIG \
--num-executors 2 \
main.py -v
```

On Amazon EMR 6.1.0 and later, to submit the job, reference the name of the Docker image. When ECR auto authentication is enabled, run the following command.

```
DOCKER_IMAGE_NAME=123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-
examples:pyspark-example
spark-submit --master yarn \
--deploy-mode cluster \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--num-executors 2 \
main.py -v
```

When the job completes, take note of the YARN application ID, and use the following command to obtain the output of the PySpark job.

```
yarn logs --applicationId application_id | grep -C2 '\[\['
LogLength:55
LogContents:
[[ 0  1  2  3  4]
 [ 5  6  7  8  9]
 [10 11 12 13 14]]
```

Using SparkR with Amazon ECR

The following example uses the SparkR Dockerfile, which will be tagged and uploaded to ECR. Once the Dockerfile is uploaded, you can run the SparkR job and refer to the Docker image from Amazon ECR.

After you launch the cluster, use SSH to connect to a core node and run the following commands to build the local Docker image from the SparkR Dockerfile example.

First, create a directory and the Dockerfile.

```
mkdir sparkr
vi sparkr/Dockerfile
```

Paste the contents of the SparkR Dockerfile and run the following commands to build a Docker image.

```
sudo docker build -t local/sparkr-example sparkr/
```

Tag and upload the locally built image to Amazon ECR, replacing *123456789123.dkr.ecr.us-east-1.amazonaws.com* with your Amazon ECR endpoint.

```
sudo docker tag local/sparkr-example 123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-examples:sparkr-example
sudo docker push 123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-examples:sparkr-example
```

Use SSH to connect to the primary node and prepare an R script with the name `sparkR.R`. Paste the following contents into the `sparkR.R` file.

```
library(SparkR)
sparkR.session(appName = "R with Spark example", sparkConfig =
 list(spark.some.config.option = "some-value"))

sqlContext <- sparkRSQL.init(spark.sparkContext)
library(randomForest)
# check release notes of randomForest
rfNews()

sparkR.session.stop()
```

On Amazon EMR 6.0.0, to submit the job, refer to the name of the Docker image. Define the additional configuration parameters to make sure that the job execution uses Docker as the runtime. When using Amazon ECR, the `YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG` must refer to the `config.json` file containing the credentials used to authenticate to ECR.

```
DOCKER_IMAGE_NAME=123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-
examples:sparkr-example
DOCKER_CLIENT_CONFIG=hdfs:///user/hadoop/config.json
spark-submit --master yarn \
--deploy-mode cluster \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG=
$DOCKER_CLIENT_CONFIG \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG=
$DOCKER_CLIENT_CONFIG \
sparkR.R
```

On Amazon EMR 6.1.0 and later, to submit the job, reference the name of the Docker image. When ECR auto authentication is enabled, run following command.

```
DOCKER_IMAGE_NAME=123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-
examples:sparkr-example
spark-submit --master yarn \
--deploy-mode cluster \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
```

```
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
sparkR.R
```

When the job has completed, note the YARN application ID, and use the following command to obtain the output of the SparkR job. This example includes testing to make sure that the randomForest library, version installed, and release notes are available.

```
yarn logs --applicationId application_id | grep -B4 -A10 "Type rfNews"
randomForest 4.6-14
Type rfNews() to see new features/changes/bug fixes.
Wishlist (formerly TODO):

* Implement the new scheme of handling classwt in classification.

* Use more compact storage of proximity matrix.

* Allow case weights by using the weights in sampling?

========================================================================
Changes in 4.6-14:
```

Use the AWS Glue Data Catalog as the metastore for Spark SQL

Using Amazon EMR release 5.8.0 or later, you can configure Spark SQL to use the AWS Glue Data Catalog as its metastore. We recommend this configuration when you require a persistent metastore or a metastore shared by different clusters, services, applications, or AWS accounts.

AWS Glue is a fully managed extract, transform, and load (ETL) service that makes it simple and cost-effective to categorize your data, clean it, enrich it, and move it reliably between various data stores. The AWS Glue Data Catalog provides a unified metadata repository across a variety of data sources and data formats, integrating with Amazon EMR as well as Amazon RDS, Amazon Redshift, Redshift Spectrum, Athena, and any application compatible with the Apache Hive metastore. AWS Glue crawlers can automatically infer schema from source data in Amazon S3 and store the associated metadata in the Data Catalog. For more information about the Data Catalog, see Populating the AWS Glue Data Catalog in the *AWS Glue Developer Guide*.

Separate charges apply for AWS Glue. There is a monthly rate for storing and accessing the metadata in the Data Catalog, an hourly rate billed per minute for AWS Glue ETL jobs and crawler runtime, and an hourly rate billed per minute for each provisioned development endpoint. The

Data Catalog allows you to store up to a million objects at no charge. If you store more than a million objects, you are charged USD$1 for each 100,000 objects over a million. An object in the Data Catalog is a table, partition, or database. For more information, see Glue Pricing.

> **Important**
>
> If you created tables using Amazon Athena or Amazon Redshift Spectrum before August 14, 2017, databases and tables are stored in an Athena-managed catalog, which is separate from the AWS Glue Data Catalog. To integrate Amazon EMR with these tables, you must upgrade to the AWS Glue Data Catalog. For more information, see Upgrading to the AWS Glue Data Catalog in the *Amazon Athena User Guide*.

Specifying AWS Glue Data Catalog as the metastore

You can specify the AWS Glue Data Catalog as the metastore using the AWS Management Console, AWS CLI, or Amazon EMR API. When you use the CLI or API, you use the configuration classification for Spark to specify the Data Catalog. In addition, with Amazon EMR 5.16.0 and later, you can use the configuration classification to specify a Data Catalog in a different AWS account. When you use the console, you can specify the Data Catalog using **Advanced Options** or **Quick Options**.

> **Note**
>
> The option to use AWS Glue Data Catalog is also available with Zeppelin because Zeppelin is installed with Spark SQL components.

Console

To specify AWS Glue Data Catalog as the Spark metastore with the new console

1. Sign in to the AWS Management Console, and open the Amazon EMR console at https://console.aws.amazon.com/emr.
2. Under **Amazon EMR on EC2** in the left navigation pane, choose **Clusters**, and then choose **Create cluster**.
3. Under **Application bundle**, choose **Spark** or **Custom**. If you customize your cluster, make sure that you select Zeppelin or Spark as one of your applications.

4. Under **AWS Glue Data Catalog settings**, select the **Use for Spark table metadata** check box.
5. Choose any other options that apply to your cluster.
6. To launch your cluster, choose **Create cluster**.

AWS CLI

To specify the AWS Glue Data Catalog as the Spark metastore with the AWS CLI

For more information about specifying a configuration classification using the AWS CLI and Amazon EMR API, see Configure applications.

- Specify the value for `hive.metastore.client.factory.class` using the `spark-hive-site` classification as shown in the following example:

```
[
  {
    "Classification": "spark-hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
 "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory"
    }
  }
]
```

To specify a Data Catalog in a different AWS account, add the `hive.metastore.glue.catalogid` property as shown in the following example. Replace *`acct-id`* with the AWS account of the Data Catalog.

```
[
  {
    "Classification": "spark-hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
 "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory",
      "hive.metastore.glue.catalogid": "acct-id"
    }
  }
]
```

IAM permissions

The EC2 instance profile for a cluster must have IAM permissions for AWS Glue actions. In addition, if you enable encryption for AWS Glue Data Catalog objects, the role must also be allowed to encrypt, decrypt and generate the AWS KMS key used for encryption.

Permissions for AWS Glue actions

If you use the default EC2 instance profile for Amazon EMR, no action is required. The `AmazonElasticMapReduceforEC2Role` managed policy that is attached to the `EMR_EC2_DefaultRole` allows all necessary AWS Glue actions. However, if you specify a custom EC2 instance profile and permissions, you must configure the appropriate AWS Glue actions. Use the `AmazonElasticMapReduceforEC2Role` managed policy as a starting point. For more information, see Service role for cluster EC2 instances (EC2 instance profile) in the *Amazon EMR Management Guide*.

Permissions for encrypting and decrypting AWS Glue Data Catalog

Your instance profile needs permission to encrypt and decrypt data using your key. You do *not* need to configure these permissions if both of the following statements apply:

- You enable encryption for AWS Glue Data Catalog objects using managed keys for AWS Glue.
- You use a cluster that's in the same AWS account as the AWS Glue Data Catalog.

Otherwise, you must add the following statement to the permissions policy attached to your EC2 instance profile.

```
[
    {
        "Version": "2012-10-17",
        "Statement": [
            {
                "Effect": "Allow",
                "Action": [
                    "kms:Decrypt",
                    "kms:Encrypt",
                    "kms:GenerateDataKey"
```

```
                ],
                "Resource": "arn:aws:kms:region:acct-
id:key/12345678-1234-1234-1234-123456789012"
            }
        ]
    }
]
```

For more information about AWS Glue Data Catalog encryption, see Encrypting your data catalog in the *AWS Glue Developer Guide*.

Resource-based permissions

If you use AWS Glue in conjunction with Hive, Spark, or Presto in Amazon EMR, AWS Glue supports resource-based policies to control access to Data Catalog resources. These resources include databases, tables, connections, and user-defined functions. For more information, see AWS Glue Resource Policies in the *AWS Glue Developer Guide*.

When using resource-based policies to limit access to AWS Glue from within Amazon EMR, the principal that you specify in the permissions policy must be the role ARN associated with the EC2 instance profile that is specified when a cluster is created. For example, for a resource-based policy attached to a catalog, you can specify the role ARN for the default service role for cluster EC2 instances, *EMR_EC2_DefaultRole* as the `Principal`, using the format shown in the following example:

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

The *acct-id* can be different from the AWS Glue account ID. This enables access from EMR clusters in different accounts. You can specify multiple principals, each from a different account.

Considerations when using AWS Glue Data Catalog

Consider the following items when using AWS Glue Data Catalog as a metastore with Spark:

- Having a default database without a location URI causes failures when you create a table. As a workaround, use the `LOCATION` clause to specify a bucket location, such as `s3://EXAMPLE-DOC-BUCKET`, when you use `CREATE TABLE`. Alternatively create tables within a database other than the default database.
- Renaming tables from within AWS Glue is not supported.

- When you create a Hive table without specifying a LOCATION, the table data is stored in the location specified by the hive.metastore.warehouse.dir property. By default, this is a location in HDFS. If another cluster needs to access the table, it fails unless it has adequate permissions to the cluster that created the table. Furthermore, because HDFS storage is transient, if the cluster terminates, the table data is lost, and the table must be recreated. We recommend that you specify a LOCATION in Amazon S3 when you create a Hive table using AWS Glue. Alternatively, you can use the hive-site configuration classification to specify a location in Amazon S3 for hive.metastore.warehouse.dir, which applies to all Hive tables. If a table is created in an HDFS location and the cluster that created it is still running, you can update the table location to Amazon S3 from within AWS Glue. For more information, see Working with Tables on the AWS Glue Console in the *AWS Glue Developer Guide*.
- Partition values containing quotes and apostrophes are not supported, for example, PARTITION (owner="Doe's").
- Column statistics are supported for emr-5.31.0 and later.
- Using Hive authorization is not supported. As an alternative, consider using AWS Glue Resource-Based Policies. For more information, see Use Resource-Based Policies for Amazon EMR Access to AWS Glue Data Catalog.

Configure Spark

You can configure Spark on Amazon EMR with configuration classifications. For more information about configuration classifications, see Configure applications.

Configuration classifications for Spark on Amazon EMR include the following:

- **spark** – Sets the maximizeResourceAllocation property to true or false. When true, Amazon EMR automatically configures spark-defaults properties based on cluster hardware configuration. For more information, see Using maximizeResourceAllocation.
- **spark-defaults** – Sets values in the spark-defaults.conf file. For more information, see Spark configuration in the Spark documentation.
- **spark-env** – Sets values in the spark-env.sh file. For more information, see Environment variables in the Spark documentation.
- **spark-hive-site** – Sets values in the hive-site.xml for Spark.
- **spark-log4j** – (Amazon EMR releases 6.7.x and lower) Sets values in the log4j.properties file. For more information, see the log4j.properties.template file on Github.

- **spark-log4j2** – (Amazon EMR releases 6.8.0 and higher) Sets values in the `log4j2.properties` file. For more information, see the log4j2.properties.template file on Github.
- **spark-metrics** – Sets values in the `metrics.properties` file. For settings and more information, see the metrics.properties.template file on Github, and Metrics in Spark documentation.

> **Note**
>
> If you're migrating Spark workloads to Amazon EMR from another platform, we recommend that you test your workloads with the Spark defaults set by Amazon EMR before you add custom configurations. Most customers see improved performance with our default settings.

Topics

Spark defaults set by Amazon EMR

The following table shows how Amazon EMR sets default values in `spark-defaults` that affect applications.

Spark defaults set by Amazon EMR

Setting	Description	Default value
`spark.executor.memory`	The amount of memory to use per executor process. For example: 1g, 2g.	This setting is determined by the core and task instance types in the cluster.
`spark.executor.cores`	The number of cores to use on each executor.	This setting is determined by the core and task instance types in the cluster.
`spark.dynamicAllocation.enabled`	When true, use dynamic resource allocation to scale the number of executors registered with an application up and down based on the workload.	`true` (with Amazon EMR 4.4.0 and higher) **Note** Spark shuffle service is automatically configured by Amazon EMR .
`spark.sql.hive.advancedPartitionPredicatePushdown.enabled`	When true, advanced partition predicate pushdown into Hive metastore is enabled.	`true`
`spark.sql.hive.stringLikePartitionPredicatePushdown.enabled`	Pushes down `startsWith` , `contains`, and `endsWith` filters into Hive metastore. **Note** Glue doesn't support predicate push down for `startsWith` , `contains`, or `endsWith`. If you are	`true`

Setting	Description	Default value
	using Glue metastore and you encounter errors due to the predicate pushdown for these functions, set this configuration to `false`.	

Configuring Spark garbage collection on Amazon EMR 6.1.0

Setting custom garbage collection configurations with `spark.driver.extraJavaOptions` and `spark.executor.extraJavaOptions` results in driver or executor launch failure with Amazon EMR 6.1 because of a conflicting garbage collection configuration with Amazon EMR 6.1.0. For Amazon EMR 6.1.0, the default garbage collection configuration is set through `spark.driver.defaultJavaOptions` and `spark.executor.defaultJavaOptions`. This configuration applies only to Amazon EMR 6.1.0. JVM options not related to garbage collection, such as those for configuring logging (`-verbose:class`), can still be set through `extraJavaOptions`. For more information, see Spark application properties.

Using `maximizeResourceAllocation`

To configure your executors to use the maximum resources possible on each node in a cluster, set `maximizeResourceAllocation` to `true` in your `spark` configuration classification. The `maximizeResourceAllocation` is specific to Amazon EMR . When you enable `maximizeResourceAllocation`, Amazon EMR calculates the maximum compute and memory resources available for an executor on an instance in the core instance group. It then sets the corresponding `spark-defaults` settings based on the calculated maximum values.

Amazon EMR calculates the maximum compute and memory resources available for an executor based on an instance type from the core instance fleet. Since each instance fleet can have different instance types and sizes within a fleet, the executor configuration that Amazon EMR uses might not be the best for your clusters, so we don't recommend using the default settings when using maximum resource allocation. Configure custom settings for your instance fleet clusters.

> **Note**
>
> You should not use the `maximizeResourceAllocation` option on clusters with other distributed applications like HBase. Amazon EMR uses custom YARN configurations for distributed applications, which can conflict with `maximizeResourceAllocation` and cause Spark applications to fail.

The following is an example Spark configuration classification with `maximizeResourceAllocation` set to `true`.

```
[
  {
    "Classification": "spark",
    "Properties": {
      "maximizeResourceAllocation": "true"
    }
  }
]
```

Settings configured in `spark-defaults` when `maximizeResourceAllocation` is enabled

Setting	Description	Value
spark.default.parallelism	Default number of partition s in RDDs returned by transformations like join, reduceByKey, and parallelize when not set by user.	2X number of CPU cores available to YARN containers.
spark.driver.memory	Amount of memory to use for the driver process, i.e. where SparkContext is initialized. (for example, 1g, 2g).	Setting is configured based on the instance types in the cluster. However, because the Spark driver application may run on either the primary or one of the core instances (for example, in YARN client and cluster modes, respectiv ely), this is set based on the

Setting	Description	Value
		smaller of the instance types in these two instance groups.
spark.executor.memory	Amount of memory to use per executor process. (for example, 1g, 2g)	Setting is configured based on the core and task instance types in the cluster.
spark.executor.cores	The number of cores to use on each executor.	Setting is configured based on the core and task instance types in the cluster.
spark.executor.instances	The number of executors.	Setting is configured based on the core and task instance types in the cluster. Set unless `spark.dynamicAllocation.enabled` explicitly set to true at the same time.

Configuring node decommissioning behavior

With Amazon EMR release 5.9.0 and higher, Spark on Amazon EMR includes a set of features to help ensure that Spark gracefully handles node termination because of a manual resize or an automatic scaling policy request. Amazon EMR implements a deny listing mechanism in Spark that is built on top of the YARN decommissioning mechanism. This mechanism helps ensure that no new tasks are scheduled on a node that is decommissioning, while at the same time allowing tasks that are already running to complete. In addition, there are features to help recover Spark jobs faster if shuffle blocks are lost when a node terminates. The recomputation process is triggered sooner and optimized to recompute faster with fewer stage retries, and jobs can be prevented from failing because of fetch failures that are caused by missing shuffle blocks.

> **Important**
>
> The `spark.decommissioning.timeout.threshold` setting was added in Amazon EMR release 5.11.0 to improve Spark resiliency when you use Spot instances. In earlier releases, when a node uses a Spot instance, and the instance is terminated because of bid price, Spark may not be able to handle the termination gracefully. Jobs may fail, and shuffle

recomputations could take a significant amount of time. For this reason, we recommend using release 5.11.0 or later if you use Spot instances.

Spark node decommissioning settings

Setting	Description	Default value
`spark.blacklist.decommissioning.enabled`	When set to `true`, Spark deny lists nodes that are in the `decommissioning` state in YARN. Spark does not schedule new tasks on executors running on that node. Tasks already running are allowed to complete.	`true`
`spark.blacklist.decommissioning.timeout`	The amount of time that a node in the `decommissioning` state is deny listed. By default, this value is set to one hour, which is also the default for `yarn.resourcemanager.decommissioning.timeout`. To ensure that a node is deny listed for its entire decommissioning period, set this value equal to or greater than `yarn.resourcemanager.decommissioning.timeout`. After the decommissioning timeout expires, the node transitions to a `decommissioned` state, and Amazon EMR can terminate the node's	`1h`

Setting	Description	Default value
	EC2 instance. If any tasks are still running after the timeout expires, they are lost or killed and rescheduled on executors running on other nodes.	
`spark.decommissioning.timeout.threshold`	Available in Amazon EMR release 5.11.0 or later. Specified in seconds. When a node transitions to the decommissioning state, if the host will decommission within a time period equal to or less than this value, Amazon EMR not only deny lists the node, but also cleans up the host state (as specified by `spark.resourceManager.cleanupExpiredHost`) without waiting for the node to transition to a decommissioned state. This allows Spark to handle Spot instance terminations better because Spot instances decommission within a 20-second timeout regardless of the value of `yarn.resourcemager.decommissioning.timeout` , which may not provide other nodes enough time to read shuffle files.	`20s`

Setting	Description	Default value
spark.resourceManager.cleanupExpiredHost	When set to true, Spark unregisters all cached data and shuffle blocks that are stored in executors on nodes that are in the decommissioned state. This speeds up the recovery process.	true
spark.stage.attempt.ignoreOnDecommissionFetchFailure	When set to true, helps prevent Spark from failing stages and eventually failing the job because of too many failed fetches from decommissioned nodes. Failed fetches of shuffle blocks from a node in the decommissioned state will not count toward the maximum number of consecutive fetch failures.	true

Spark ThriftServer environment variable

Spark sets the Hive Thrift Server Port environment variable, HIVE_SERVER2_THRIFT_PORT, to 10001.

Changing Spark default settings

You change the defaults in spark-defaults.conf using the spark-defaults configuration classification or the maximizeResourceAllocation setting in the spark configuration classification.

The following procedures show how to modify settings using the CLI or console.

To create a cluster with spark.executor.memory set to 2g using the CLI

- Create a cluster with Spark installed and `spark.executor.memory` set to 2g, using the following command, which references a file, `myConfig.json` stored in Amazon S3.

```
aws emr create-cluster --release-label emr-7.2.0 --applications Name=Spark \
--instance-type m5.xlarge --instance-count 2 --service-role EMR_DefaultRole_V2
 --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole --configurations https://
s3.amazonaws.com/mybucket/myfolder/myConfig.json
```

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

`myConfig.json`:

```
[
    {
      "Classification": "spark-defaults",
      "Properties": {
        "spark.executor.memory": "2G"
      }
    }
  ]
```

To create a cluster with spark.executor.memory set to 2g using the console

1. Navigate to the new Amazon EMR console and select **Switch to the old console** from the side navigation. For more information on what to expect when you switch to the old console, see Using the old console.
2. Choose **Create cluster**, **Go to advanced options**.
3. Choose **Spark**.
4. Under **Edit software settings**, leave **Enter configuration** selected and enter the following configuration:

```
classification=spark-defaults,properties=[spark.executor.memory=2G]
```

5. Select other options, choose and then choose **Create cluster**.

To set maximizeResourceAllocation

- Create a cluster with Spark installed and `maximizeResourceAllocation` set to true using the AWS CLI, referencing a file, `myConfig.json`, stored in Amazon S3.

```
aws emr create-cluster --release-label emr-7.2.0 --applications Name=Spark \
--instance-type m5.xlarge --instance-count 2 --service-role EMR_DefaultRole_V2
 --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole --configurations https://
s3.amazonaws.com/mybucket/myfolder/myConfig.json
```

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

`myConfig.json`:

```
[
  {
    "Classification": "spark",
    "Properties": {
      "maximizeResourceAllocation": "true"
    }
  }
]
```

> **Note**
>
> With Amazon EMR version 5.21.0 and later, you can override cluster configurations and specify additional configuration classifications for each instance group in a running cluster. You do this by using the Amazon EMR console, the AWS Command Line Interface (AWS

CLI), or the AWS SDK. For more information, see Supplying a Configuration for an Instance Group in a Running Cluster.

Migrating from Apache Log4j 1.x to Log4j 2.x

Apache Spark releases 3.2.x and earlier use the legacy Apache Log4j 1.x and the `log4j.properties` file to configure Log4j in Spark processes. Apache Spark releases 3.3.0 and later use Apache Log4j 2.x and the `log4j2.properties` file to configure Log4j in Spark processes.

If you have configured Apache Spark Log4j using an Amazon EMR release lower than 6.8.0, then you must remove the legacy `spark-log4j` configuration classification and migrate to the `spark-log4j2` configuration classification and key format before you can upgrade to Amazon EMR 6.8.0 or later. The legacy `spark-log4j` classification causes cluster creation to fail with a `ValidationException` error in Amazon EMR releases 6.8.0 and later. You will not be charged for a failure related to the Log4j incompatibility, but you must remove the defunct `spark-log4j` configuration classification to continue.

For more information about migrating from Apache Log4j 1.x to Log4j 2.x, see the Apache Log4j Migration Guide and the Spark Log4j 2 Template on Github.

> **Note**
>
> With Amazon EMR , Apache Spark uses a `log4j2.properties` file rather than the .xml file described in the Apache Log4j Migration Guide. Also, we do not recommend using the Log4j 1.x bridge method to convert to Log4j 2.x.

Optimize Spark performance

Amazon EMR provides multiple performance optimization features for Spark. This topic explains each optimization feature in detail.

For more information on how to set Spark configuration, see Configure Spark.

Adaptive query execution

Adaptive query execution is a framework for reoptimizing query plans based on runtime statistics. Starting with Amazon EMR 5.30.0, the following adaptive query execution optimizations from Apache Spark 3 are available on Apache Amazon EMR Runtime for Spark 2.

- Adaptive join conversion
- Adaptive coalescing of shuffle partitions

Adaptive Join Conversion

Adaptive join conversion improves query performance by converting sort-merge-join operations to broadcast-hash-joins operations based on runtime sizes of query stages. Broadcast-hash-joins tend to perform better when one side of the join is small enough to efficiently broadcast its output across all executors, thereby avoiding the need to shuffle exchange and sort both sides of the join. Adaptive join conversion broadens the range of cases when Spark automatically performs broadcast-hash-joins.

This feature is enabled by default. It can be disabled by setting `spark.sql.adaptive.enabled` to `false`, which also disables the adaptive query execution framework. Spark decides to convert a sort-merge-join to a broadcast-hash-join when the runtime size statistic of one of the join sides does not exceed `spark.sql.autoBroadcastJoinThreshold`, which defaults to 10,485,760 bytes (10 MiB).

Adaptive Coalescing of Shuffle Partitions

Adaptive coalescing of shuffle partitions improves query performance by coalescing small contiguous shuffle partitions to avoid the overhead of having too many small tasks. This allows you to configure a higher number of initial shuffle partitions upfront that then gets reduced at runtime to a targeted size, improving the chances of having more evenly distributed shuffle partitions.

This feature is enabled by default unless `spark.sql.shuffle.partitions` is explicitly set. It can be enabled by setting `spark.sql.adaptive.coalescePartitions.enabled` to `true`. Both the initial number of shuffle partitions and target partition size can be tuned using the `spark.sql.adaptive.coalescePartitions.minPartitionNum` and `spark.sql.adaptive.advisoryPartitionSizeInBytes` properties respectively. See the following table for further details on the related Spark properties for this feature.

Spark adaptive coalesce partition properties

Property	Default value	Description
`spark.sql.adaptive.coalescePartitions.enabled`	true, unless `spark.sql.shuffle.partitions` is explicitly set	When true and spark.sql.adaptive.enabled is true, Spark coalesces contiguous shuffle partitions according to the target size (specified by `spark.sql.adaptive.advisoryPartitionSizeInBytes`), to avoid too many small tasks.
`spark.sql.adaptive.advisoryPartitionSizeInBytes`	64MB	The advisory size in bytes of the shuffle partition when coalescing. This configuration only has an effect when `spark.sql.adaptive.enabled` and `spark.sql.adaptive.coalescePartitions.enabled` are both `true`.
`spark.sql.adaptive.coalescePartitions.minPartitionNum`	25	The minimum number of shuffle partitions after coalescing. This configuration only has an effect when `spark.sql.adaptive.enabled` and `spark.sql.adaptive.coalescePartitions.enabled` are both `true`.
`spark.sql.adaptive.coalescePartitions.initialPartitionNum`	1000	The initial number of shuffle partitions before coalescing. This configuration only has an effect when `spark.sql`

Property	Default value	Description
		.adaptive.enabled and spark.sql.adaptive .coalescePartition s.enabled are both true.

Dynamic partition pruning

Dynamic partition pruning improves job performance by more accurately selecting the specific partitions within a table that need to be read and processed for a specific query. By reducing the amount of data read and processed, significant time is saved in job execution. With Amazon EMR 5.26.0, this feature is enabled by default. With Amazon EMR 5.24.0 and 5.25.0, you can enable this feature by setting the Spark property `spark.sql.dynamicPartitionPruning.enabled` from within Spark or when creating clusters.

Spark dynamic partition pruning partition properties

Property	Default Value	Description
`spark.sql.dynamicPartitionPruning.enabled`	`true`	When true, enable dynamic partition pruning.
`spark.sql.optimizer.dynamicPartitionPruning.enforceBroadcastReuse`	`true`	When `true`, Spark performs a defensive check before query execution to ensure that reuse of broadcast exchanges in dynamic pruning filters is not broken by later preparation rules, such as user-defined columnar rules. When reuse is broken and this config is `true`, Spark removes the affected dynamic pruning filters to guard against performance and correctness issues. Correctness issues

Property	Default Value	Description
		may arise when the broadcast exchange of the dynamic pruning filter yields different , inconsistent results from the broadcast exchange of the corresponding join operation . Setting this configuration to `false` should be done with caution; it allows working around scenarios, such as when reuse is broken by user-defined columnar rules. When Adaptive Query Execution is enabled, broadcast reuse is always enforced.

This optimization improves upon the existing capabilities of Spark 2.4.2, which only supports pushing down static predicates that can be resolved at plan time.

The following are examples of static predicate push down in Spark 2.4.2.

```
partition_col = 5

partition_col IN (1,3,5)

partition_col between 1 and 3

partition_col = 1 + 3
```

Dynamic partition pruning allows the Spark engine to dynamically infer at runtime which partitions need to be read and which can be safely eliminated. For example, the following query involves two tables: the `store_sales` table that contains all of the total sales for all stores and is partitioned by region, and the `store_regions` table that contains a mapping of regions for each country. The tables contain data about stores distributed around the globe, but we are only querying data for North America.

```
select ss.quarter, ss.region, ss.store, ss.total_sales
from store_sales ss, store_regions sr
where ss.region = sr.region and sr.country = 'North America'
```

Without dynamic partition pruning, this query will read all regions before filtering out the subset of regions that match the results of the subquery. With dynamic partition pruning, this query will read and process only the partitions for the regions returned in the subquery. This saves time and resources by reading less data from storage and processing less records.

Flattening scalar subqueries

This optimization improves the performance of queries that have scalar subqueries over the same table. With Amazon EMR 5.26.0, this feature is enabled by default. With Amazon EMR 5.24.0 and 5.25.0, you can enable it by setting the Spark property `spark.sql.optimizer.flattenScalarSubqueriesWithAggregates.enabled` from within Spark or when creating clusters. When this property is set to true, the query optimizer flattens aggregate scalar subqueries that use the same relation if possible. The scalar subqueries are flattened by pushing any predicates present in the subquery into the aggregate functions and then performing one aggregation, with all the aggregate functions, per relation.

Following is a sample of a query that will benefit from this optimization.

```
select (select avg(age) from students                    /* Subquery 1 */
                 where age between 5 and 10) as group1,
       (select avg(age) from students                    /* Subquery 2 */
                 where age between 10 and 15) as group2,
       (select avg(age) from students                    /* Subquery 3 */
                 where age between 15 and 20) as group3
```

The optimization rewrites the previous query as:

```
select c1 as group1, c2 as group2, c3 as group3
from (select avg (if(age between 5 and 10, age, null)) as c1,
             avg (if(age between 10 and 15, age, null)) as c2,
             avg (if(age between 15 and 20, age, null)) as c3 from students);
```

Notice that the rewritten query reads the student table only once, and the predicates of the three subqueries are pushed into the `avg` function.

DISTINCT before INTERSECT

This optimization optimizes joins when using INTERSECT. With Amazon EMR 5.26.0, this feature is enabled by default. With Amazon EMR 5.24.0 and 5.25.0, you can enable it by setting the Spark property `spark.sql.optimizer.distinctBeforeIntersect.enabled` from within Spark or when creating clusters. Queries using INTERSECT are automatically converted to use a left-semi join. When this property is set to true, the query optimizer pushes the DISTINCT operator to the children of INTERSECT if it detects that the DISTINCT operator can make the left-semi join a BroadcastHashJoin instead of a SortMergeJoin.

Following is a sample of a query that will benefit from this optimization.

```
(select item.brand brand from store_sales, item
     where store_sales.item_id = item.item_id)
intersect
(select item.brand cs_brand from catalog_sales, item
     where catalog_sales.item_id = item.item_id)
```

Without enabling this property `spark.sql.optimizer.distinctBeforeIntersect.enabled`, the query will be rewritten as follows.

```
select distinct brand from
  (select item.brand brand from store_sales, item
     where store_sales.item_id = item.item_id)
left semi join
   (select item.brand cs_brand from catalog_sales, item
     where catalog_sales.item_id = item.item_id)
 on brand <=> cs_brand
```

When you enable this property `spark.sql.optimizer.distinctBeforeIntersect.enabled`, the query will be rewritten as follows.

```
select brand from
  (select distinct item.brand brand from store_sales, item
     where store_sales.item_id = item.item_id)
left semi join
   (select distinct item.brand cs_brand from catalog_sales, item
     where catalog_sales.item_id = item.item_id)
```

```
on brand <=> cs_brand
```

Bloom filter join

This optimization can improve the performance of some joins by pre-filtering one side of a join using a [Bloom filter](https://en.wikipedia.org/wiki/Bloom_filter) generated from the values from the other side of the join. With Amazon EMR 5.26.0, this feature is enabled by default. With Amazon EMR 5.25.0, you can enable this feature by setting the Spark property `spark.sql.bloomFilterJoin.enabled` to `true` from within Spark or when creating clusters.

The following is an example query that can benefit from a Bloom filter.

```
select count(*)
from sales, item
where sales.item_id = item.id
and item.category in (1, 10, 16)
```

When this feature is enabled, the Bloom filter is built from all item ids whose category is in the set of categories being queried. While scanning the sales table, the Bloom filter is used to determine which sales are for items that are definitely not in the set defined by the Bloom filter. Thus these identified sales can be filtered out as early as possible.

Optimized join reorder

This optimization can improve query performance by reordering joins involving tables with filters. With Amazon EMR 5.26.0, this feature is enabled by default. With Amazon EMR 5.25.0, you can enable this feature by setting the Spark configuration parameter `spark.sql.optimizer.sizeBasedJoinReorder.enabled` to true. The default behavior in Spark is to join tables from left to right, as listed in the query. This strategy can miss opportunities to execute smaller joins with filters first, in order to benefit more expensive joins later.

The example query below reports all returned items from all stores in a country. Without optimized join reorder, Spark joins the two large tables `store_sales` and `store_returns` first, and then joins them with `store` and eventually with `item`.

```
select ss.item_value, sr.return_date, s.name, i.desc,
from store_sales ss, store_returns sr, store s, item i
where ss.id = sr.id and ss.store_id = s.id and ss.item_id = i.id
```

```
and s.country = 'USA'
```

With optimized join reorder, Spark joins `store_sales` with `store` first since `store` has a filter and is smaller than `store_returns` and `broadcastable`. Then Spark joins with `store_returns` and finally with `item`. If `item` had a filter and was broadcastable, it would also qualify for reorder, resulting in `store_sales` joining with `store`, then `item`, and eventually with `store_returns`.

Spark Result Fragment Caching

Amazon EMR 6.6.0 and higher include the optional Spark Result Fragment Caching feature that automatically caches result fragments. These result fragments are parts of results from subtrees of queries that are stored in an Amazon S3 bucket of your choosing. The stored query result fragments are reused on subsequent query executions, resulting in faster queries.

Result Fragment Caching analyzes your Spark SQL queries, and caches eligible result fragments in your specified S3 location. On subsequent query runs, the usable query result fragments are automatically detected and fetched from S3. Result Fragment Caching differs from Result Set Caching, where subsequent queries have to exactly match the original query to return results from the cache. When used for queries that repeatedly target a static subset of your data, result fragment caching speeds performance significantly.

Consider the following query, which counts orders until the year 2022:

```
select
    l_returnflag,
    l_linestatus,
    count(*) as count_order
from
    lineitem
where
    l_shipdate <= current_date
    and year(l_shipdate) == '2022'
group by
    l_returnflag,
    l_linestatus
```

As time progresses, this query needs to run every day to report the total sales for the year. Without Result Fragment Caching, the results for all days of the year will need to be recomputed every day.

The query will get slower over time and will be slowest at the end of the year, when all 365 days of results will need to be recomputed.

When you activate Result Fragment Caching, you use results for the all previous days of the year from the cache. Each day, the feature must recompute only one day of results. After the feature computes the result fragment, the feature caches the fragment. As a result, cache-enabled query times are fast and they remain constant for each subsequent query.

Enabling Spark Result Fragment Caching

To enable Spark Result Fragment Caching, perform the following steps:

1. Create a cache bucket in Amazon S3 and authorize read/write access for EMRFS. For more information, see Authorizing access to EMRFS data in Amazon S3.
2. Set Amazon EMR Spark config to enable the feature.

```
spark.subResultCache.enabled = true
spark.subResultCache.fs.root.path = s3://DOC-EXAMPLE-BUCKET/cache_dir/
```

3. Enable S3 lifecycle management for the bucket to automatically clean cache files.
4. Optionally, configure the reductionRationThreshold and maxBufferSize properties to further tune the feature.

```
spark.sql.subResultCache.reductionRatioThreshold
spark.sql.subResultCache.maxBufferSize
```

Considerations when using Result Fragment Caching

The cost savings when you use results already cached in Amazon S3 rather than recompute them grows with the number of times the same cached results can be used. Queries with large table scans followed by filters or hash aggregations that reduce the result size by factor of at least 8 (that is, a ratio of at least 8:1 in input size:results) will benefit most from this feature. The greater the reduction ratio between the input and the results, the greater is the cost benefit. Queries with smaller reduction ratios, but that contain expensive computation steps between the table scan and filter or aggregations will also benefit, as long as the cost to produce the results is greater than the cost to fetch them from Amazon S3. By default, Result Fragment Caching takes effect only when it detects that a reduction ratio will be at least 8:1.

When your queries repeatedly reuse cached results, the benefits of this feature are greatest. Rolling and incremental window queries are good examples. For instance, a 30-day rolling window query that has already run for 29 days, would only need to pull 1/30th of the target data from its original input source and would use cached result fragments for the 29 previous days. An incremental window query would benefit even more, since the start of the window remains fixed: on every invocation of the query, a smaller percentage of the processing will require reading from the input source.

The following are additional considerations when using Result Fragment Caching:

- Queries that don't target the same data with the same query fragments will have a low cache hit rate, hence will not benefit from this feature.
- Queries with low reduction ratios that do not contain expensive computation steps will result in cached results that are roughly as expensive to read as they were to initially process.
- The first query will always demonstrate a minor regression due to the cost of writing to cache.
- The Result Fragment Caching feature works exclusively with Parquet files. Other file formats are not supported.
- The Result Fragment Caching feature buffers will only attempt to cache scans with file split sizes of 128 MB or larger. With the default Spark configuration, Result Fragment Caching will be disabled if the scan size (total size across all files being scanned) divided by the number of executor cores is less than 128 MB. When any of the Spark configurations listed below are set, the file split size will be:

  ```
  min(maxPartitionBytes, max(openCostInBytes, scan size / minPartitionNum))
  ```

  - spark.sql.leafNodeDefaultParallelism (default value is spark.default.parallelism)
  - spark.sql.files.minPartitionNum (default value is spark.sql.leafNodeDefaultParallelism)
  - spark.sql.files.openCostInBytes
  - spark.sql.files.maxPartitionBytes
- The Result Fragment Caching feature caches at the RDD partition granularity. The previously described reduction ratio that defaults to 8:1 is assessed per RDD partition. Workloads with per-RDD reduction ratios both greater and less than 8:1 may see smaller performance benefits than workloads with per-RDD reduction ratios that are consistently less than 8:1.
- The Result Fragment Caching feature uses a 16MB write buffer by default for each RDD partition being cached.If more than 16mb will be cached per RDD partition, the cost of determining that a write is not possible may result in a performance regression.

- While, by default, Result Fragment Caching will not attempt to cache RDD partition results with a reduction ratio smaller than 8:1 and will cap its write buffer at 16MB, both of these values are tunable through the following configurations:

```
spark.sql.subResultCache.reductionRatioThreshold (default: 8.0)
spark.sql.subResultCache.maxBufferSize (default: 16MB, max: 64MB)
```

- Multiple clusters using the same Amazon EMR release can share the same cache location. To ensure result correctness, Result Fragment Caching will not use cache results written by different releases of Amazon EMR .
- Result Fragment Caching will be disabled automatically for Spark Streaming use cases or when RecordServer, Apache Ranger, or AWS Lake Formation is used.
- The result fragment cache read/writes use EMRFS and Amazon S3 buckets. CSE/ SSE S3/ SSE KMS encryption are supported.

Use the Nvidia RAPIDS Accelerator for Apache Spark

With Amazon EMR release 6.2.0 and later, you can use the RAPIDS Accelerator for Apache Spark plugin by Nvidia to accelerate Spark using EC2 graphics processing unit (GPU) instance types. RAPIDS Accelerator will GPU-accelerate your Apache Spark 3.0 data science pipelines without code changes, and speed up data processing and model training while substantially lowering infrastructure costs.

The following sections guide you through configuring your EMR cluster to use the Spark-RAPIDS Plugin for Spark.

Choose instance types

To use the Nvidia Spark-RAPIDS plugin for Spark, the core and task instance groups must use EC2 GPU instance types that meet the Hardware requirements of Spark-RAPIDS. To view a complete list of Amazon EMR supported GPU instance types, please see Supported instance types in the *Amazon EMR Management Guide*. Instance type for the primary instance group can be either GPU or non-GPU types, but ARM instance types aren't supported.

Set up application configurations for your cluster

1. Enable Amazon EMR to install the plugins on your new cluster

To install plugins, supply the following configuration when you create your cluster:

```
{
 "Classification":"spark",
 "Properties":{
  "enableSparkRapids":"true"
 }
}
```

2. Configure YARN to use GPU

For details on how to use GPU on YARN, see Using GPU on YARN in Apache Hadoop documentation. The following examples show sample YARN configurations for Amazon EMR 6.x and 7.x releases:

Amazon EMR 7.x

Example YARN configuration for Amazon EMR 7.x

```
{
    "Classification":"yarn-site",
    "Properties":{
        "yarn.nodemanager.resource-plugins":"yarn.io/gpu",
        "yarn.resource-types":"yarn.io/gpu",
        "yarn.nodemanager.resource-plugins.gpu.allowed-gpu-devices":"auto",
        "yarn.nodemanager.resource-plugins.gpu.path-to-discovery-executables":"/usr/
bin",
        "yarn.nodemanager.linux-container-executor.cgroups.mount":"true",
        "yarn.nodemanager.linux-container-executor.cgroups.mount-path":"/spark-
rapids-cgroup",
        "yarn.nodemanager.linux-container-executor.cgroups.hierarchy":"yarn",
        "yarn.nodemanager.container-
executor.class":"org.apache.hadoop.yarn.server.nodemanager.LinuxContainerExecutor"
    }
},{
    "Classification":"container-executor",
    "Properties":{

    },
    "Configurations":[
        {
            "Classification":"gpu",
            "Properties":{
```

```
                "module.enabled":"true"
            }
        },
        {
            "Classification":"cgroups",
            "Properties":{
                "root":"/spark-rapids-cgroup",
                "yarn-hierarchy":"yarn"
            }
        }
    ]
}
```

Amazon EMR 6.x

Example YARN configuration for Amazon EMR 6.x

```
{
    "Classification":"yarn-site",
    "Properties":{
        "yarn.nodemanager.resource-plugins":"yarn.io/gpu",
        "yarn.resource-types":"yarn.io/gpu",
        "yarn.nodemanager.resource-plugins.gpu.allowed-gpu-devices":"auto",
        "yarn.nodemanager.resource-plugins.gpu.path-to-discovery-executables":"/usr/
bin",
        "yarn.nodemanager.linux-container-executor.cgroups.mount":"true",
        "yarn.nodemanager.linux-container-executor.cgroups.mount-path":"/sys/fs/
cgroup",
        "yarn.nodemanager.linux-container-executor.cgroups.hierarchy":"yarn",
        "yarn.nodemanager.container-
executor.class":"org.apache.hadoop.yarn.server.nodemanager.LinuxContainerExecutor"
    }
},{
    "Classification":"container-executor",
    "Properties":{

    },
    "Configurations":[
        {
            "Classification":"gpu",
            "Properties":{
                "module.enabled":"true"
            }
        },
```

```
        {
            "Classification":"cgroups",
            "Properties":{
                "root":"/sys/fs/cgroup",
                "yarn-hierarchy":"yarn"
            }
        }
    ]
}
```

3. Configure Spark to use RAPIDS

Here are the required configurations to enable Spark to use RAPIDS plugin:

```
{
 "Classification":"spark-defaults",
 "Properties":{
  "spark.plugins":"com.nvidia.spark.SQLPlugin",
  "spark.executor.resource.gpu.discoveryScript":"/usr/lib/spark/scripts/gpu/
getGpusResources.sh",
  "spark.executor.extraLibraryPath":"/usr/local/cuda/targets/x86_64-linux/lib:/usr/
local/cuda/extras/CUPTI/lib64:/usr/local/cuda/compat/lib:/usr/local/cuda/lib:/usr/
local/cuda/lib64:/usr/lib/hadoop/lib/native:/usr/lib/hadoop-lzo/lib/native:/docker/usr/
lib/hadoop/lib/native:/docker/usr/lib/hadoop-lzo/lib/native"
 }
}
```

XGBoost4J-Spark library in XGBoost documentation is also available when the Spark RAPIDS plugin is enabled on your cluster. You can use the following configuration to integrate XGBoost with you Spark job:

```
{
 "Classification":"spark-defaults",
 "Properties":{
  "spark.submit.pyFiles":"/usr/lib/spark/jars/xgboost4j-spark_3.0-1.4.2-0.3.0.jar"
 }
}
```

For additional Spark configurations that you can use to tune a GPU-accelerated EMR cluster, please refer to the Rapids Accelerator for Apache Spark tuning guide in Nvidia.github.io documentation.

4. Configure YARN Capacity Scheduler

`DominantResourceCalculator` must be configured to enable GPU scheduling and isolation. For more information, please refer to Using GPU on YARN in Apache Hadoop documentation.

```
{
 "Classification":"capacity-scheduler",
 "Properties":{
  "yarn.scheduler.capacity.resource-
calculator":"org.apache.hadoop.yarn.util.resource.DominantResourceCalculator"
 }
}
```

5. Create a JSON file to include your configurations

You can create a JSON file that contains your configuration to use the RAPIDS plugin for your Spark cluster. You supply the file later when you launch your cluster.

You can store the file locally or on S3. For more information of how to supply application configurations for your clusters, see Configure applications.

Use the following sample files as templates to build your own configurations.

Amazon EMR 7.x

Example `my-configurations.json` file for Amazon EMR 7.x

```
[
    {
        "Classification":"spark",
        "Properties":{
            "enableSparkRapids":"true"
        }
    },
    {
        "Classification":"yarn-site",
        "Properties":{
            "yarn.nodemanager.resource-plugins":"yarn.io/gpu",
            "yarn.resource-types":"yarn.io/gpu",
            "yarn.nodemanager.resource-plugins.gpu.allowed-gpu-devices":"auto",
            "yarn.nodemanager.resource-plugins.gpu.path-to-discovery-executables":"/
usr/bin",
            "yarn.nodemanager.linux-container-executor.cgroups.mount":"true",
```

```
            "yarn.nodemanager.linux-container-executor.cgroups.mount-path":"/spark-
rapids-cgroup",
            "yarn.nodemanager.linux-container-executor.cgroups.hierarchy":"yarn",
            "yarn.nodemanager.container-
executor.class":"org.apache.hadoop.yarn.server.nodemanager.LinuxContainerExecutor"
        }
    },
    {
        "Classification":"container-executor",
        "Properties":{

        },
        "Configurations":[
            {
                "Classification":"gpu",
                "Properties":{
                    "module.enabled":"true"
                }
            },
            {
                "Classification":"cgroups",
                "Properties":{
                    "root":"/spark-rapids-cgroup",
                    "yarn-hierarchy":"yarn"
                }
            }
        ]
    },
    {
        "Classification":"spark-defaults",
        "Properties":{
            "spark.plugins":"com.nvidia.spark.SQLPlugin",
            "spark.executor.resource.gpu.discoveryScript":"/usr/lib/spark/scripts/
gpu/getGpusResources.sh",
            "spark.executor.extraLibraryPath":"/usr/local/cuda/targets/x86_64-
linux/lib:/usr/local/cuda/extras/CUPTI/lib64:/usr/local/cuda/compat/lib:/usr/local/
cuda/lib:/usr/local/cuda/lib64:/usr/lib/hadoop/lib/native:/usr/lib/hadoop-lzo/lib/
native:/docker/usr/lib/hadoop/lib/native:/docker/usr/lib/hadoop-lzo/lib/native",
            "spark.submit.pyFiles":"/usr/lib/spark/jars/xgboost4j-
spark_3.0-1.4.2-0.3.0.jar",
            "spark.rapids.sql.concurrentGpuTasks":"1",
            "spark.executor.resource.gpu.amount":"1",
            "spark.executor.cores":"2",
            "spark.task.cpus":"1",
```

```
            "spark.task.resource.gpu.amount":"0.5",
            "spark.rapids.memory.pinnedPool.size":"0",
            "spark.executor.memoryOverhead":"2G",
            "spark.locality.wait":"0s",
            "spark.sql.shuffle.partitions":"200",
            "spark.sql.files.maxPartitionBytes":"512m"
        }
    },
    {
        "Classification":"capacity-scheduler",
        "Properties":{
            "yarn.scheduler.capacity.resource-
calculator":"org.apache.hadoop.yarn.util.resource.DominantResourceCalculator"
        }
    }
]
```

Amazon EMR 6.x

Example `my-configurations.json` file for Amazon EMR 6.x

```
[
    {
        "Classification":"spark",
        "Properties":{
            "enableSparkRapids":"true"
        }
    },
    {
        "Classification":"yarn-site",
        "Properties":{
            "yarn.nodemanager.resource-plugins":"yarn.io/gpu",
            "yarn.resource-types":"yarn.io/gpu",
            "yarn.nodemanager.resource-plugins.gpu.allowed-gpu-devices":"auto",
            "yarn.nodemanager.resource-plugins.gpu.path-to-discovery-executables":"/
usr/bin",
            "yarn.nodemanager.linux-container-executor.cgroups.mount":"true",
            "yarn.nodemanager.linux-container-executor.cgroups.mount-path":"/sys/fs/
cgroup",
            "yarn.nodemanager.linux-container-executor.cgroups.hierarchy":"yarn",
            "yarn.nodemanager.container-
executor.class":"org.apache.hadoop.yarn.server.nodemanager.LinuxContainerExecutor"
        }
    },
```

```
    {
        "Classification":"container-executor",
        "Properties":{

        },
        "Configurations":[
            {
                "Classification":"gpu",
                "Properties":{
                    "module.enabled":"true"
                }
            },
            {
                "Classification":"cgroups",
                "Properties":{
                    "root":"/sys/fs/cgroup",
                    "yarn-hierarchy":"yarn"
                }
            }
        ]
    },
    {
        "Classification":"spark-defaults",
        "Properties":{
            "spark.plugins":"com.nvidia.spark.SQLPlugin",
            "spark.executor.resource.gpu.discoveryScript":"/usr/lib/spark/scripts/
gpu/getGpusResources.sh",
            "spark.executor.extraLibraryPath":"/usr/local/cuda/targets/x86_64-
linux/lib:/usr/local/cuda/extras/CUPTI/lib64:/usr/local/cuda/compat/lib:/usr/local/
cuda/lib:/usr/local/cuda/lib64:/usr/lib/hadoop/lib/native:/usr/lib/hadoop-lzo/lib/
native:/docker/usr/lib/hadoop/lib/native:/docker/usr/lib/hadoop-lzo/lib/native",
            "spark.submit.pyFiles":"/usr/lib/spark/jars/xgboost4j-
spark_3.0-1.4.2-0.3.0.jar",
            "spark.rapids.sql.concurrentGpuTasks":"1",
            "spark.executor.resource.gpu.amount":"1",
            "spark.executor.cores":"2",
            "spark.task.cpus":"1",
            "spark.task.resource.gpu.amount":"0.5",
            "spark.rapids.memory.pinnedPool.size":"0",
            "spark.executor.memoryOverhead":"2G",
            "spark.locality.wait":"0s",
            "spark.sql.shuffle.partitions":"200",
            "spark.sql.files.maxPartitionBytes":"512m"
        }
```

```
    },
    {
        "Classification":"capacity-scheduler",
        "Properties":{
            "yarn.scheduler.capacity.resource-
calculator":"org.apache.hadoop.yarn.util.resource.DominantResourceCalculator"
        }
    }
]
```

Add a bootstrap action for your cluster

For more information on how to supply bootstrap action scripts when you create your cluster, see Bootstrap action basics in the *Amazon EMR Management Guide*.

The following example scripts show how to make a bootstrap action file for Amazon EMR 6.x and 7.x:

Amazon EMR 7.x

Example `my-bootstrap-action.sh` file for Amazon EMR 7.x

To use YARN to manage GPU resources with Amazon EMR 7.x releases, you must manually mount CGroup v1 on your cluster. You can do this with as bootstrap action script, as shown in this example.

```
#!/bin/bash
set -ex

sudo mkdir -p /spark-rapids-cgroup/devices
sudo mount -t cgroup -o devices cgroupv1-devices /spark-rapids-cgroup/devices
sudo chmod a+rwx -R /spark-rapids-cgroup
```

Amazon EMR 6.x

Example `my-bootstrap-action.sh` file for Amazon EMR 6.x

For Amazon EMR 6.x releases, you must open CGroup permissions to YARN on your cluster. You can do this with a bootstrap action script, as shown in this example.

```
#!/bin/bash
set -ex
```

```
sudo chmod a+rwx -R /sys/fs/cgroup/cpu,cpuacct
sudo chmod a+rwx -R /sys/fs/cgroup/devices
```

Launch your cluster

The last step is to launch your cluster with the cluster configurations mentioned above. Here's an example command to launch a cluster from the Amazon EMR CLI:

```
 aws emr create-cluster \
--release-label emr-7.2.0 \
--applications Name=Hadoop Name=Spark \
--service-role EMR_DefaultRole_V2 \
--ec2-attributes KeyName=my-key-pair,InstanceProfile=EMR_EC2_DefaultRole \
--instance-groups InstanceGroupType=MASTER,InstanceCount=1,InstanceType=m4.4xlarge \

                  InstanceGroupType=CORE,InstanceCount=1,InstanceType=g4dn.2xlarge \

                  InstanceGroupType=TASK,InstanceCount=1,InstanceType=g4dn.2xlarge \
--configurations file:///my-configurations.json \
--bootstrap-actions Name='My Spark Rapids Bootstrap action',Path=s3://my-bucket/my-bootstrap-action.sh
```

Access the Spark shell

The Spark shell is based on the Scala REPL (Read-Eval-Print-Loop). It allows you to create Spark programs interactively and submit work to the framework. You can access the Spark shell by connecting to the primary node with SSH and invoking `spark-shell`. For more information about connecting to the primary node, see Connect to the primary node using SSH in the *Amazon EMR Management Guide*. The following examples use Apache HTTP Server access logs stored in Amazon S3.

> **Note**
>
> The bucket in these examples is available to clients that can access US East (N. Virginia).

By default, the Spark shell creates its own SparkContext object called `sc`. You can use this context if it is required within the REPL. sqlContext is also available in the shell and it is a HiveContext.

Example Use the Spark shell to count the occurrences of a string in a file stored in Amazon S3

This example uses `sc` to read a text file that's stored in Amazon S3.

```
scala> sc
res0: org.apache.spark.SparkContext = org.apache.spark.SparkContext@404721db

scala> val textFile = sc.textFile("s3://elasticmapreduce/samples/hive-ads/tables/
impressions/dt=2009-04-13-08-05/ec2-0-51-75-39.amazon.com-2009-04-13-08-05.log")
```

Spark creates the textFile and associated [data structure](). Next, the example counts the number of lines in the log file with the string "cartoonnetwork.com":

```
scala> val linesWithCartoonNetwork = textFile.filter(line =>
 line.contains("cartoonnetwork.com")).count()
linesWithCartoonNetwork: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[2] at
 filter at <console>:23
<snip>
<Spark program runs>
scala> linesWithCartoonNetwork
res2: Long = 9
```

Example Use the Python-based Spark shell to count the occurrences of a string in a file stored in Amazon S3

Spark also includes a Python-based shell, `pyspark`, that you can use to prototype Spark programs written in Python. Just as with `spark-shell`, invoke `pyspark` on the primary node; it also has the same [SparkContext]() object.

```
>>> sc
<pyspark.context.SparkContext object at 0x7fe7e659fa50>
>>> textfile = sc.textFile("s3://elasticmapreduce/samples/hive-ads/tables/impressions/
dt=2009-04-13-08-05/ec2-0-51-75-39.amazon.com-2009-04-13-08-05.log")
```

Spark creates the textFile and associated [data structure](). Next, the example counts the number of lines in the log file with the string "cartoonnetwork.com".

```
>>> linesWithCartoonNetwork = textfile.filter(lambda line: "cartoonnetwork.com" in
 line).count()
15/06/04 17:12:22 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library from the
 embedded binaries
```

```
15/06/04 17:12:22 INFO lzo.LzoCodec: Successfully loaded & initialized native-lzo
 library [hadoop-lzo rev EXAMPLE]
15/06/04 17:12:23 INFO fs.EmrFileSystem: Consistency disabled, using
 com.amazon.ws.emr.hadoop.fs.s3n.S3NativeFileSystem as filesystem implementation
<snip>
<Spark program continues>
>>> linesWithCartoonNetwork
9
```

Use Amazon SageMaker Spark for machine learning

When using Amazon EMR release 5.11.0 and later, the `aws-sagemaker-spark-sdk` component is installed along with Spark. This component installs Amazon SageMaker Spark and associated dependencies for Spark integration with Amazon SageMaker. You can use Amazon SageMaker Spark to construct Spark machine learning (ML) pipelines using Amazon SageMaker stages. For more information, see the Amazon SageMaker Spark README on GitHub and Using Apache Spark with Amazon SageMaker in the *Amazon SageMaker Developer Guide*.

Write a Spark application

Spark applications can be written in Scala, Java, or Python. There are several examples of Spark applications located on Spark examples topic in the Apache Spark documentation. The Estimating Pi example is shown below in the three natively supported applications. You can also view complete examples in `$SPARK_HOME/examples` and at GitHub. For more information about how to build JARs for Spark, see the Quick start topic in the Apache Spark documentation.

Scala

To avoid Scala compatibility issues, we suggest you use Spark dependencies for the correct Scala version when you compile a Spark application for an Amazon EMR cluster. The Scala version you should use depends on the version of Spark installed on your cluster. For example, Amazon EMR release 5.30.1 uses Spark 2.4.5, which is built with Scala 2.11. If your cluster uses Amazon EMR release 5.30.1, use Spark dependencies for Scala 2.11. For more information about the Scala versions used by Spark, see the Apache Spark documentation.

```
package org.apache.spark.examples
import scala.math.random
import org.apache.spark._
```

```
/** Computes an approximation to pi */
object SparkPi {
  def main(args: Array[String]) {
    val conf = new SparkConf().setAppName("Spark Pi")
    val spark = new SparkContext(conf)
    val slices = if (args.length > 0) args(0).toInt else 2
    val n = math.min(100000L * slices, Int.MaxValue).toInt // avoid overflow
    val count = spark.parallelize(1 until n, slices).map { i =>
      val x = random * 2 - 1
      val y = random * 2 - 1
      if (x*x + y*y < 1) 1 else 0
    }.reduce(_ + _)
    println("Pi is roughly " + 4.0 * count / n)
    spark.stop()
  }
}
```

Java

```
package org.apache.spark.examples;

import org.apache.spark.SparkConf;
import org.apache.spark.api.java.JavaRDD;
import org.apache.spark.api.java.JavaSparkContext;
import org.apache.spark.api.java.function.Function;
import org.apache.spark.api.java.function.Function2;

import java.util.ArrayList;
import java.util.List;

/**
 * Computes an approximation to pi
 * Usage: JavaSparkPi [slices]
 */
public final class JavaSparkPi {

  public static void main(String[] args) throws Exception {
    SparkConf sparkConf = new SparkConf().setAppName("JavaSparkPi");
    JavaSparkContext jsc = new JavaSparkContext(sparkConf);

    int slices = (args.length == 1) ? Integer.parseInt(args[0]) : 2;
    int n = 100000 * slices;
    List<Integer> l = new ArrayList<Integer>(n);
```

```
    for (int i = 0; i < n; i++) {
      l.add(i);
    }

    JavaRDD<Integer> dataSet = jsc.parallelize(l, slices);

    int count = dataSet.map(new Function<Integer, Integer>() {
      @Override
      public Integer call(Integer integer) {
        double x = Math.random() * 2 - 1;
        double y = Math.random() * 2 - 1;
        return (x * x + y * y < 1) ? 1 : 0;
      }
    }).reduce(new Function2<Integer, Integer, Integer>() {
      @Override
      public Integer call(Integer integer, Integer integer2) {
        return integer + integer2;
      }
    });

    System.out.println("Pi is roughly " + 4.0 * count / n);

    jsc.stop();
  }
}
```

Python

```
import argparse
import logging
from operator import add
from random import random

from pyspark.sql import SparkSession

logger = logging.getLogger(__name__)
logging.basicConfig(level=logging.INFO, format="%(levelname)s: %(message)s")


def calculate_pi(partitions, output_uri):
    """
    Calculates pi by testing a large number of random numbers against a unit circle
```

```
    inscribed inside a square. The trials are partitioned so they can be run in
    parallel on cluster instances.

    :param partitions: The number of partitions to use for the calculation.
    :param output_uri: The URI where the output is written, typically an Amazon S3
                       bucket, such as 's3://example-bucket/pi-calc'.
    """

    def calculate_hit(_):
        x = random() * 2 - 1
        y = random() * 2 - 1
        return 1 if x**2 + y**2 < 1 else 0

    tries = 100000 * partitions
    logger.info(
        "Calculating pi with a total of %s tries in %s partitions.", tries, partitions
    )
    with SparkSession.builder.appName("My PyPi").getOrCreate() as spark:
        hits = (
            spark.sparkContext.parallelize(range(tries), partitions)
            .map(calculate_hit)
            .reduce(add)
        )
        pi = 4.0 * hits / tries
        logger.info("%s tries and %s hits gives pi estimate of %s.", tries, hits, pi)
        if output_uri is not None:
            df = spark.createDataFrame([(tries, hits, pi)], ["tries", "hits", "pi"])
            df.write.mode("overwrite").json(output_uri)


if __name__ == "__main__":
    parser = argparse.ArgumentParser()
    parser.add_argument(
        "--partitions",
        default=2,
        type=int,
        help="The number of parallel partitions to use when calculating pi.",
    )
    parser.add_argument(
        "--output_uri", help="The URI where output is saved, typically an S3 bucket."
    )
    args = parser.parse_args()

    calculate_pi(args.partitions, args.output_uri)
```

Improve Spark performance with Amazon S3

Amazon EMR offers features to help optimize performance when using Spark to query, read and write data saved in Amazon S3.

S3 Select can improve query performance for CSV and JSON files in some applications by "pushing down" processing to Amazon S3.

The EMRFS S3-optimized committer is an alternative to the OutputCommitter class, which uses the multipart uploads feature of EMRFS to improve performance when writing Parquet files to Amazon S3 using Spark SQL, DataFrames, and Datasets.

Topics

Use S3 Select with Spark to improve query performance

With Amazon EMR release 5.17.0 and later, you can use S3 Select with Spark on Amazon EMR . *S3 Select* allows applications to retrieve only a subset of data from an object. For Amazon EMR , the computational work of filtering large data sets for processing is "pushed down" from the cluster to Amazon S3, which can improve performance in some applications and reduces the amount of data transferred between Amazon EMR and Amazon S3.

S3 Select is supported with CSV and JSON files using `s3selectCSV` and `s3selectJSON` values to specify the data format. For more information and examples, see Specify S3 Select in your code.

Is S3 Select right for my application?

We recommend that you benchmark your applications with and without S3 Select to see if using it may be suitable for your application.

Use the following guidelines to determine if your application is a candidate for using S3 Select:

- Your query filters out more than half of the original data set.
- Your network connection between Amazon S3 and the Amazon EMR cluster has good transfer speed and available bandwidth. Amazon S3 does not compress HTTP responses, so the response size is likely to increase for compressed input files.

Considerations and limitations

- Amazon S3 server-side encryption with customer-provided encryption keys (SSE-C) and client-side encryption are not supported.
- The `AllowQuotedRecordDelimiters` property is not supported. If this property is specified, the query fails.
- Only CSV and JSON files in UTF-8 format are supported. Multi-line CSVs are not supported.
- Only uncompressed or gzip files are supported.
- Spark CSV and JSON options such as `nanValue`, `positiveInf`, `negativeInf`, and options related to corrupt records (for example, failfast and dropmalformed mode) are not supported.
- Using commas (,) within decimals is not supported. For example, `10,000` is not supported and `10000` is.
- Comment characters in the last line are not supported.
- Empty lines at the end of a file are not processed.
- The following filters are not pushed down to Amazon S3:
  - Aggregate functions such as `COUNT()` and `SUM()`.
  - Filters that `CAST()` an attribute. For example, `CAST(stringColumn as INT) = 1`.
  - Filters with an attribute that is an object or is complex. For example, `intArray[1] = 1, objectColumn.objectNumber = 1`.
  - Filters for which the value is not a literal value. For example, `intColumn1 = intColumn2`
  - Only S3 Select supported data types are supported with the documented limitations.

Specify S3 Select in your code

The following examples demonstrate how to specify S3 Select for CSV using Scala, SQL, R, and PySpark. You can use S3 Select for JSON in the same way. For a listing of options, their default values, and limitations, see Options.

PySpark

```
spark
  .read
  .format("s3selectCSV") // "s3selectJson" for Json
  .schema(...) // optional, but recommended
  .options(...) // optional
  .load("s3://path/to/my/datafiles")
```

R

```
read.df("s3://path/to/my/datafiles", "s3selectCSV", schema, header = "true",
 delimiter = "\t")
```

Scala

```
spark
  .read
  .format("s3selectCSV") // "s3selectJson" for Json
  .schema(...) // optional, but recommended
  .options(...) // optional. Examples:
  // .options(Map("quote" -> "\'", "header" -> "true")) or
  // .option("quote", "\'").option("header", "true")
  .load("s3://path/to/my/datafiles")
```

SQL

```
CREATE TEMPORARY VIEW MyView (number INT, name STRING) USING s3selectCSV OPTIONS
 (path "s3://path/to/my/datafiles", header "true", delimiter "\t")
```

Options

The following options are available when using `s3selectCSV` and `s3selectJSON`. If not specified, default values are used.

Options with S3selectCSV

Option	Default	Usage
`compression`	`"none"`	Indicates whether compressi on is used. `"gzip"` is the only setting supported besides `"none"`.
`delimiter`	`","`	Specifies the field delimiter.
`quote`	`'\"'`	Specifies the quote character . Specifying an empty string is not supported and results in a malformed XML error.
`escape`	`'\\'`	Specifies the escape character .
`header`	`"false"`	`"false"` specifies that there is no header. `"true"` specifies that a header is in the first line. Only headers in the first line are supported, and empty lines before a header are not supported.
comment	`"#"`	Specifies the comment character. The comment indicator cannot be disabled. In other words, a value of `\u0000` is not supported.
`nullValue`	""	

Options with S3selectJSON

Option	Default	Usage
`compression`	`"none"`	Indicates whether compression is used. `"gzip"` is the only setting supported besides `"none"`.
`multiline`	"false"	`"false"` specifies that the JSON is in S3 Select `LINES` format, meaning that each line in the input data contains a single JSON object. `"true"` specifies that the JSON is in S3 Select `DOCUMENT` format, meaning that a JSON object can span multiple lines in the input data.

Use the EMRFS S3-optimized committer

The EMRFS S3-optimized committer is an alternative OutputCommitter implementation that is optimized for writing files to Amazon S3 when using EMRFS. The EMRFS S3-optimized committer improves application performance by avoiding list and rename operations done in Amazon S3 during job and task commit phases. The committer is available with Amazon EMR release 5.19.0 and later, and is enabled by default with Amazon EMR 5.20.0 and later. The committer is used for Spark jobs that use Spark SQL, DataFrames, or Datasets. Starting with Amazon EMR 6.4.0, this committer can be used for all common formats including parquet, ORC, and text-based formats (including CSV and JSON). For releases prior to Amazon EMR 6.4.0, only the Parquet format is supported. There are circumstances under which the committer is not used. For more information, see Requirements for the EMRFS S3-optimized committer.

Topics

- Job tuning considerations
- Enable the EMRFS S3-optimized committer for Amazon EMR 5.19.0

Requirements for the EMRFS S3-optimized committer

The EMRFS S3-optimized committer is used when the following conditions are met:

- You run Spark jobs that use Spark SQL, DataFrames, or Datasets to write files to Amazon S3. Starting with Amazon EMR 6.4.0, this committer can be used for all common formats including parquet, ORC, and text-based formats (including CSV and JSON). For releases prior to Amazon EMR 6.4.0, only the Parquet format is supported.
- Multipart uploads are enabled in Amazon EMR . This is the default. For more information, see The EMRFS S3-optimized committer and multipart uploads.
- Spark's built-in file format support is used. The built-in file format support is used in the following circumstances:
  - For Hive metastore tables, when `spark.sql.hive.convertMetastoreParquet` is set to `true` for Parquet tables, or `spark.sql.hive.convertMetastoreOrc` is set to `true` for Orc tables with Amazon EMR 6.4.0 or above. These are the default settings.
  - When jobs write to file format data sources or tables—for example, the target table is created with the `USING parquet` clause.
  - When jobs write to non-partitioned Hive metastore Parquet tables. Spark's built-in Parquet support does not support partitioned Hive tables, which is a known limitation. For more information, see Hive metastore Parquet table conversion in the Apache Spark SQL, DataFrames and Datasets Guide.
- Spark job operations that write to a default partition location—for example, `${table_location}/k1=v1/k2=v2/`—use the committer. The committer is not used if a job operation writes to a custom partition location—for example, if a custom partition location is set using the `ALTER TABLE SQL` command.
- The following values for Spark must be used:
  - The `spark.sql.parquet.fs.optimized.committer.optimization-enabled` property must be set to `true`. This is the default setting with Amazon EMR 5.20.0 and later. With Amazon EMR 5.19.0, the default value is `false`. For information about configuring this value, see Enable the EMRFS S3-optimized committer for Amazon EMR 5.19.0.
  - If writing to non-partitioned Hive metastore tables, only Parquet and Orc file formats are supported. `spark.sql.hive.convertMetastoreParquet` must be set to `true` if writing to

non-partitioned Parquet Hive metastore tables. `spark.sql.hive.convertMetastoreOrc` must be set to `true` if writing to non-partitioned Orc Hive metastore tables. These are the default settings.

- `spark.sql.parquet.output.committer.class` must be set to `com.amazon.emr.committer.EmrOptimizedSparkSqlParquetOutputCommitter`. This is the default setting.
- `spark.sql.sources.commitProtocolClass` must be set to `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol` or `org.apache.spark.sql.execution.datasources.SQLHadoopMapReduceCommitProtocol`. `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol` is the default setting for the Amazon EMR 5.x series version 5.30.0 and higher, and for the Amazon EMR 6.x series version 6.2.0 and higher. `org.apache.spark.sql.execution.datasources.SQLHadoopMapReduceCommitProtocol` is the default setting for previous Amazon EMR versions.
- If Spark jobs overwrite partitioned Parquet datasets with dynamic partition columns, then the `partitionOverwriteMode` write option and `spark.sql.sources.partitionOverwriteMode` must be set to `static`. This is the default setting.

> **Note**
>
> The `partitionOverwriteMode` write option was introduced in Spark 2.4.0. For Spark version 2.3.2, included with Amazon EMR release 5.19.0, set the `spark.sql.sources.partitionOverwriteMode` property.

Occasions when EMRFS S3-optimized committer is not used

Generally, the EMRFS S3-optimized committer isn't used in the following situations.

Situation	Why the committer is not used
When you write to HDFS	The committer only supports writing to Amazon S3 using EMRFS.
When you use the S3A file system	The committer only supports EMRFS.

Situation	Why the committer is not used
When you use MapReduce or Spark's RDD API	The committer only supports using SparkSQL, DataFrame, or Dataset APIs.

The following Scala examples demonstrate some additional situations that prevent the EMRFS S3-optimized committer from being used in whole (the first example) and in part (the second example).

Example – Dynamic partition overwrite mode

The following Scala example instructs Spark to use a different commit algorithm, which prevents use of the EMRFS S3-optimized committer altogether. The code sets the `partitionOverwriteMode` property to `dynamic` to overwrite only those partitions to which you're writing data. Then, dynamic partition columns are specified by `partitionBy`, and the write mode is set to `overwrite`.

```
val dataset = spark.range(0, 10)
  .withColumn("dt", expr("date_sub(current_date(), id)"))

dataset.write.mode("overwrite")
  .option("partitionOverwriteMode", "dynamic")
  .partitionBy("dt")
  .parquet("s3://EXAMPLE-DOC-BUCKET/output")
```

You must configure all three settings to avoid using the EMRFS S3-optimized committer. When you do so, Spark executes a different commit algorithm that's specified in Spark's commit protocol. For Amazon EMR 5.x releases earlier than 5.30.0 and for Amazon EMR 6.x releases earlier than 6.2.0, the commit protocol uses Spark's staging directory, which is a temporary directory created under the output location that starts with `.spark-staging`. The algorithm sequentially renames partition directories, which can negatively impact performance. For more information about Amazon EMR releases 5.30.0 and later and 6.2.0 and later, see Use the EMRFS S3-optimized commit protocol.

The algorithm in Spark 2.4.0 follows these steps:

1. Task attempts write their output to partition directories under Spark's staging directory—for example, `${outputLocation}/spark-staging-${jobID}/k1=v1/k2=v2/`.

2. For each partition written, the task attempt keeps track of relative partition paths—for example, `k1=v1/k2=v2`.
3. When a task completes successfully, it provides the driver with all relative partition paths that it tracked.
4. After all tasks complete, the job commit phase collects all the partition directories that successful task attempts wrote under Spark's staging directory. Spark sequentially renames each of these directories to its final output location using directory tree rename operations.
5. The staging directory is deleted before the job commit phase completes.

Example – Custom partition location

In this example, the Scala code inserts into two partitions. One partition has a custom partition location. The other partition uses the default partition location. The EMRFS S3-optimized committer is only used for writing task output to the partition that uses the default partition location.

```
val table = "dataset"
val location = "s3://bucket/table"

spark.sql(s"""
  CREATE TABLE $table (id bigint, dt date)
  USING PARQUET PARTITIONED BY (dt)
  LOCATION '$location'
""")

// Add a partition using a custom location
val customPartitionLocation = "s3://bucket/custom"
spark.sql(s"""
  ALTER TABLE $table ADD PARTITION (dt='2019-01-28')
  LOCATION '$customPartitionLocation'
""")

// Add another partition using default location
spark.sql(s"ALTER TABLE $table ADD PARTITION (dt='2019-01-29')")

def asDate(text: String) = lit(text).cast("date")

spark.range(0, 10)
  .withColumn("dt",
```

```
    when($"id" > 4, asDate("2019-01-28")).otherwise(asDate("2019-01-29")))
  .write.insertInto(table)
```

The Scala code creates the following Amazon S3 objects:

```
custom/part-00001-035a2a9c-4a09-4917-8819-e77134342402.c000.snappy.parquet
custom_$folder$
table/_SUCCESS
table/dt=2019-01-29/part-00000-035a2a9c-4a09-4917-8819-e77134342402.c000.snappy.parquet
table/dt=2019-01-29_$folder$
table_$folder$
```

When writing to partitions at custom locations, Spark uses a commit algorithm similar to the previous example, which is outlined below. As with the earlier example, the algorithm results in sequential renames, which may negatively impact performance.

1. When writing output to a partition at a custom location, tasks write to a file under Spark's staging directory, which is created under the final output location. The name of the file includes a random UUID to protect against file collisions. The task attempt keeps track of each file along with the final desired output path.
2. When a task completes successfully, it provides the driver with the files and their final desired output paths.
3. After all tasks complete, the job commit phase sequentially renames all files that were written for partitions at custom locations to their final output paths.
4. The staging directory is deleted before the job commit phase completes.

The EMRFS S3-optimized committer and multipart uploads

To use the EMRFS S3-optimized committer, you must enable multipart uploads for Amazon EMR . Multipart uploads are enabled by default. You can re-enable it if required. For more information, see Configure multipart upload for Amazon S3 in the *Amazon EMR Management Guide*.

The EMRFS S3-optimized committer uses the transaction-like characteristics of multipart uploads to ensure files written by task attempts only appear in the job's output location upon task commit. By using multipart uploads in this way, the committer improves task commit performance over the default FileOutputCommitter algorithm version 2. When using the EMRFS S3-optimized committer, there are some key differences from traditional multipart upload behavior to consider:

- Multipart uploads are always performed regardless of the file size. This differs from the default behavior of EMRFS, where the `fs.s3n.multipart.uploads.split.size` property controls the file size at which multipart uploads are triggered.
- Multipart uploads are left in an incomplete state for a longer period of time until the task commits or aborts. This differs from the default behavior of EMRFS where a multipart upload completes when a task finishes writing a given file.

Because of these differences, if a Spark Executor JVM crashes or is killed while tasks are running and writing data to Amazon S3, incomplete multipart uploads are more likely to be left behind. For this reason, when you use the EMRFS S3-optimized committer, be sure to follow the best practices for managing failed multipart uploads. For more information, see Best practices for working with Amazon S3 buckets in the *Amazon EMR Management Guide*.

Job tuning considerations

The EMRFS S3-optimized committer consumes a small amount of memory for each file written by a task attempt until the task gets committed or aborted. In most jobs, the amount of memory consumed is negligible. For jobs that have long-running tasks that write a large number of files, the memory that the committer consumes may be noticeable and require adjustments to the memory allocated for Spark executors. You can tune executor memory using the `spark.executor.memory` property. As a guideline, a single task writing 100,000 files would typically require an additional 100MB of memory. For more information, see Application properties in the Apache Spark Configuration documentation.

Enable the EMRFS S3-optimized committer for Amazon EMR 5.19.0

If you are using Amazon EMR 5.19.0 , you can manually set the `spark.sql.parquet.fs.optimized.committer.optimization-enabled` property to `true` when you create a cluster or from within Spark if you are using Amazon EMR .

Enabling the EMRFS S3-optimized committer when creating a cluster

Use the `spark-defaults` configuration classification to set the `spark.sql.parquet.fs.optimized.committer.optimization-enabled` property to `true`. For more information, see Configure applications.

Enabling the EMRFS S3-optimized committer from Spark

You can set `spark.sql.parquet.fs.optimized.committer.optimization-enabled` to `true` by hard-coding it in a `SparkConf`, passing it as a `--conf` parameter in the Spark shell or `spark-submit` and `spark-sql` tools, or in `conf/spark-defaults.conf`. For more information, see Spark configuration in Apache Spark documentation.

The following example shows how to enable the committer while running a spark-sql command.

```
spark-sql \
  --conf spark.sql.parquet.fs.optimized.committer.optimization-enabled=true \
  -e "INSERT OVERWRITE TABLE target_table SELECT * FROM source_table;"
```

Use the EMRFS S3-optimized commit protocol

The EMRFS S3-optimized commit protocol is an alternative FileCommitProtocol implementation that is optimized for writing files with Spark dynamic partition overwrite to Amazon S3 when using EMRFS. The protocol improves application performance by avoiding rename operations in Amazon S3 during the Spark dynamic partition overwrite job commit phase.

Note that the Use the EMRFS S3-optimized committer also improves performance by avoiding rename operations. However, it doesn't work for dynamic partition overwrite cases, while the commit protocol's improvements only target dynamic partition overwrite cases.

The commit protocol is available with Amazon EMR release 5.30.0 and later and 6.2.0 and later and is enabled by default. Amazon EMR added a parallelism improvement starting with release 5.31.0. The protocol is used for Spark jobs that use Spark SQL, DataFrames, or Datasets. There are circumstances under which the commit protocol is not used. For more information, see Requirements for the EMRFS S3-optimized commit protocol.

Topics

Requirements for the EMRFS S3-optimized commit protocol

The EMRFS S3-optimized commit protocol is used when the following conditions are met:

- You run Spark jobs that use Spark SQL, DataFrames, or Datasets to overwrite partitioned tables.
- You run Spark jobs whose partition overwrite mode is `dynamic`.
- Multipart uploads are enabled in Amazon EMR . This is the default. For more information, see The EMRFS S3-optimized commit protocol and multipart uploads.
- The filesystem cache for EMRFS is enabled. This is the default. Check that the setting `fs.s3.impl.disable.cache` is set to `false`.
- Spark's built-in data source support is used. Built-in data source support is used in the following circumstances:
  - When jobs write to built-in data sources or tables.
  - When jobs write to the Hive metastore Parquet table. This happens when `spark.sql.hive.convertInsertingPartitionedTable` and `spark.sql.hive.convertMetastoreParquet` are both set to true. These are the default settings.
  - When jobs write to the Hive metastore ORC table. This happens when `spark.sql.hive.convertInsertingPartitionedTable` and `spark.sql.hive.convertMetastoreOrc` are both set to `true`. These are the default settings.
- Spark job operations that write to a default partition location – for example, `${table_location}/k1=v1/k2=v2/` – use the commit protocol. The protocol is not used if a job operation writes to a custom partition location – for example, if a custom partition location is set using the `ALTER TABLE SQL` command.
- The following values for Spark must be used:
  - `spark.sql.sources.commitProtocolClass` must be set to `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol`. This is the default setting for Amazon EMR releases 5.30.0 and higher, and 6.2.0 and higher.
  - The `partitionOverwriteMode` write option or `spark.sql.sources.partitionOverwriteMode` must be set to `dynamic`. The default setting is `static`.

> **Note**
>
> The `partitionOverwriteMode` write option was introduced in Spark 2.4.0. For Spark version 2.3.2, included with Amazon EMR release 5.19.0, set the `spark.sql.sources.partitionOverwriteMode` property.

- If Spark jobs overwrite to the Hive metastore Parquet table, `spark.sql.hive.convertMetastoreParquet`, `spark.sql.hive.convertInsertingPartitionedTable`, and `spark.sql.hive.convertMetastore.partitionOverwriteMode` must be set to `true`. There are the default settings.
- If Spark jobs overwrite to the Hive metastore ORC table, `spark.sql.hive.convertMetastoreOrc`, `spark.sql.hive.convertInsertingPartitionedTable`, and `spark.sql.hive.convertMetastore.partitionOverwriteMode` must be set to `true`. There are the default settings.

Example – Dynamic partition overwrite mode

In this Scala example, optimization is triggered. First, you set the `partitionOverwriteMode` property to `dynamic`. This only overwrites those partitions to which you're writing data. Then, you specify dynamic partition columns with `partitionBy` and set the write mode to `overwrite`.

```
val dataset = spark.range(0, 10)
  .withColumn("dt", expr("date_sub(current_date(), id)"))

dataset.write.mode("overwrite")                       // "overwrite" instead of "insert"
  .option("partitionOverwriteMode", "dynamic")  // "dynamic" instead of "static"
  .partitionBy("dt")                            // partitioned data instead of
 unpartitioned data
  .parquet("s3://EXAMPLE-DOC-BUCKET/output")    // "s3://" to use Amazon EMR file
 system, instead of "s3a://" or "hdfs://"
```

When the EMRFS S3-optimized commit protocol is not used

Generally, the EMRFS S3-optimized commit protocol works the same as open source default Spark SQL commit protocol, `org.apache.spark.sql.execution.datasources.SQLHadoopMapReduceCommitProtocol`. Optimization won't occur in the following situations.

Situation	Why the commit protocol is not used
When you write to HDFS	The commit protocol only supports writing to Amazon S3 using EMRFS.

Situation	Why the commit protocol is not used
When you use the S3A file system	The commit protocol only supports EMRFS.
When you use MapReduce or Spark's RDD API	The commit protocol only supports using SparkSQL, DataFrame, or Dataset APIs.
When the dynamic partition overwrite isn't triggered	The commit protocol only optimizes dynamic partition overwrite cases. For other cases, see Use the EMRFS S3-optimized committer.

The following Scala examples demonstrate some additional situations that the EMRFS S3-optimized commit protocol delegates to `SQLHadoopMapReduceCommitProtocol`.

Example – Dynamic partition overwrite mode with custom partition location

In this example, the Scala programs overwrites two partitions in dynamic partition overwrite mode. One partition has a custom partition location. The other partition uses the default partition location. The EMRFS S3-optimized commit protocol only improves the partition that uses the default partition location.

```
val table = "dataset"
val inputView = "tempView"
val location = "s3://bucket/table"

spark.sql(s"""
  CREATE TABLE $table (id bigint, dt date)
  USING PARQUET PARTITIONED BY (dt)
  LOCATION '$location'
""")

// Add a partition using a custom location
val customPartitionLocation = "s3://bucket/custom"
spark.sql(s"""
  ALTER TABLE $table ADD PARTITION (dt='2019-01-28')
  LOCATION '$customPartitionLocation'
""")

// Add another partition using default location
spark.sql(s"ALTER TABLE $table ADD PARTITION (dt='2019-01-29')")
```

```
def asDate(text: String) = lit(text).cast("date")

spark.range(0, 10)
  .withColumn("dt",
    when($"id" > 4, asDate("2019-01-28")).otherwise(asDate("2019-01-29")))
  .createTempView(inputView)

// Set partition overwrite mode to 'dynamic'
spark.sql(s"SET spark.sql.sources.partitionOverwriteMode=dynamic")

spark.sql(s"INSERT OVERWRITE TABLE $table SELECT * FROM $inputView")
```

The Scala code creates the following Amazon S3 objects:

```
custom/part-00001-035a2a9c-4a09-4917-8819-e77134342402.c000.snappy.parquet
custom_$folder$
table/_SUCCESS
table/dt=2019-01-29/part-00000-035a2a9c-4a09-4917-8819-e77134342402.c000.snappy.parquet
table/dt=2019-01-29_$folder$
table_$folder$
```

Note

Writing to custom partition locations in earlier Spark versions may result in data loss. In this example, partition `dt='2019-01-28'` would be lost. For more details, see SPARK-35106. This is fixed in Amazon EMR release 5.33.0 and later, excluding 6.0.x and 6.1.x.

When writing to partitions at custom locations, Spark uses a commit algorithm similar to the previous example, which is outlined below. As with the earlier example, the algorithm results in sequential renames, which may negatively impact performance.

The algorithm in Spark 2.4.0 follows these steps:

1. When writing output to a partition at a custom location, tasks write to a file under Spark's staging directory, which is created under the final output location. The name of the file includes a random UUID to protect against file collisions. The task attempt keeps track of each file along with the final desired output path.
2. When a task completes successfully, it provides the driver with the files and their final desired output paths.

3. After all tasks complete, the job commit phase sequentially renames all files that were written for partitions at custom locations to their final output paths.
4. The staging directory is deleted before the job commit phase completes.

The EMRFS S3-optimized commit protocol and multipart uploads

To use make use of the optimization for dynamic partition overwrite in the EMRFS S3-optimized commit protocol, multipart uploads must be enabled in Amazon EMR . Multipart uploads are enabled by default. You can re-enable it if required. For more information, see Configure multipart upload for Amazon S3 in the *Amazon EMR Management Guide*.

During dynamic partition overwrite, the EMRFS S3-optimized commit protocol uses the transaction-like characteristics of multipart uploads to ensure files written by task attempts only appear in the job's output location upon job commit. By using multipart uploads in this way, the commit protocol improves job commit performance over the default `SQLHadoopMapReduceCommitProtocol`. When using the EMRFS S3-optimized commit protocol, there are some key differences from traditional multipart upload behavior to consider:

- Multipart uploads are always performed regardless of the file size. This differs from the default behavior of EMRFS, where the `fs.s3n.multipart.uploads.split.size` property controls the file size at which multipart uploads are triggered.
- Multipart uploads are left in an incomplete state for a longer period of time until the task commits or aborts. This differs from the default behavior of EMRFS where a multipart upload completes when a task finishes writing a given file.

Because of these differences, if a Spark Executor JVM crashes or is killed while tasks are running and writing data to Amazon S3, or a Spark Driver JVM crashes or is killed while a job is running, incomplete multipart uploads are more likely to be left behind. For this reason, when you use the EMRFS S3-optimized commit protocol, be sure to follow the best practices for managing failed multipart uploads. For more information, see Best practices for working with Amazon S3 buckets in the *Amazon EMR Management Guide*.

Job tuning considerations

On Spark executors, the EMRFS S3-optimized commit protocol consumes a small amount of memory for each file written by a task attempt until the task gets committed or aborted. In most jobs, the amount of memory consumed is negligible.

On Spark drivers, the EMRFS S3-optimized commit protocol requires memory to store metadata info of each committed file until the job gets committed or aborted. In most jobs, default Spark driver memory setting is negligible.

For jobs that have long-running tasks that write a large number of files, the memory that the commit protocol consumes may be noticeable and require adjustments to the memory allocated for Spark, especially for Spark executors. You can tune memory using the `spark.driver.memory` property for Spark drivers, and the `spark.executor.memory` property for Spark executors. As a guideline, a single task writing 100,000 files would typically require an additional 100MB of memory. For more information, see [Application properties] in the Apache Spark Configuration documentation.

Retry Amazon S3 requests with EMRFS

This topic provides information about the retry strategies that you can use when making requests to Amazon S3 with EMRFS. When your request rate increases, S3 tries to scale to support the new rate. During this process S3 can throttle requests and return a `503 Slow Down` error. To improve the success rate of your S3 requests, you can adjust your retry strategy by configuring properties in your `emrfs-site` configuration.

You can adjust your retry strategy in the following ways.

- Increase the maximum retry limit for the default exponential backoff retry strategy.
- Enable and configure the additive-increase/multiplicative-decrease (AIMD) retry strategy. AIMD is supported for Amazon EMR releases 6.4.0 and later.

Use the default exponential backoff strategy

By default, EMRFS uses an exponential backoff strategy to retry Amazon S3 requests. The default EMRFS retry limit is 15. To avoid an S3 `503 Slow Down` error, you can increase the retry limit when you create a new cluster, on a running cluster, or at application runtime.

To increase the retry limit, you must change the value for `fs.s3.maxRetries` in your `emrfs-site` configuration. The following example configuration sets `fs.s3.maxRetries` to a custom value of 30.

```
[
    {
      "Classification": "emrfs-site",
```

```
      "Properties": {
        "fs.s3.maxRetries": "30"
      }
    }
]
```

For more information about working with configuration objects, see Configure applications.

Use the AIMD retry strategy

With Amazon EMR release 6.4.0 and later, EMRFS supports an alternative retry strategy based on an additive-increase/multiplicative-decrease (AIMD) model. The AIMD retry strategy is especially useful when you work with large Amazon EMR clusters.

AIMD calculates a custom request rate using data about recent successful requests. This strategy decreases the number of throttled requests and the total attempts required per request.

To enable the AIMD retry strategy, you must set the `fs.s3.aimd.enabled` property to `true` in your `emrfs-site` configuration as in the following example.

```
[
    {
      "Classification": "emrfs-site",
      "Properties": {
        "fs.s3.aimd.enabled": "true"
      }
    }
]
```

For more information about working with configuration objects, see Configure applications.

Advanced AIMD retry settings

You can configure the properties listed in the following table to refine retry behavior when you use the AIMD retry strategy. For most use cases, we recommend that you use the default values.

Advanced AIMD retry strategy properties

Property	Default value	Description
`fs.s3.aimd.increaseIncrement`	0.1	Controls how quickly the request rate increases when

Property	Default value	Description
		consecutive requests are successful.
`fs.s3.aimd.reductionFactor`	2	Controls how quickly the request rate decreases when Amazon S3 returns a 503 response. The default factor of 2 cuts the request rate in half.
`fs.s3.aimd.minRate`	0.1	Sets the lower bound for the request rate when requests experience sustained throttling by S3.
`fs.s3.aimd.initialRate`	5500	Sets the initial request rate, which then changes according to the values that you specify for `fs.s3.aimd.increaseIncrement` and `fs.s3.aimd.reductionFactor` . The initial rate is also used for GET requests, and scaled proportionally (3500/5500) for PUT requests.
`fs.s3.aimd.adjustWindow`	2	Controls how frequently the request rate is adjusted, measured in number of responses.
`fs.s3.aimd.maxAttempts`	100	Sets the maximum number of attempts to try a request.

Add a Spark step

You can use Amazon EMR steps to submit work to the Spark framework installed on an EMR cluster. For more information, see Steps in the Amazon EMR Management Guide. In the console and CLI, you do this using a Spark application step, which runs the `spark-submit` script as a step on your behalf. With the API, you use a step to invoke `spark-submit` using `command-runner.jar`.

For more information about submitting applications to Spark, see the Submitting applications topic in the Apache Spark documentation.

To submit a Spark step using the console

1. Open the Amazon EMR console at https://console.aws.amazon.com/emr.
2. In the **Cluster List**, choose the name of your cluster.
3. Scroll to the **Steps** section and expand it, then choose **Add step**.
4. In the **Add Step** dialog box:
   - For **Step type**, choose **Spark application**.
   - For **Name**, accept the default name (Spark application) or type a new name.
   - For **Deploy mode**, choose **Client** or **Cluster** mode. Client mode launches the driver program on the cluster's primary instance, while cluster mode launches your driver program on the cluster. For client mode, the driver's log output appears in the step logs, while for cluster mode, the driver's log output appears in the logs for the first YARN container. For more information, see Cluster mode overview in the Apache Spark documentation.
   - Specify the desired **Spark-submit options**. For more information about `spark-submit` options, see Launching applications with spark-submit.
   - For **Application location**, specify the local or S3 URI path of the application.
   - For **Arguments**, leave the field blank.
   - For **Action on failure**, accept the default option (**Continue**).
5. Choose **Add**. The step appears in the console with a status of Pending.
6. The status of the step changes from **Pending** to **Running** to **Completed** as the step runs. To update the status, choose the **Refresh** icon above the **Actions** column.

7. The results of the step are located in the Amazon EMR console Cluster Details page next to your step under **Log Files** if you have logging configured. You can optionally find step information in the log bucket you configured when you launched the cluster.

To submit work to Spark using the AWS CLI

Submit a step when you create the cluster or use the `aws emr add-steps` subcommand in an existing cluster.

1. Use `create-cluster` as shown in the following example.

   Note

   Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --name "Add Spark Step Cluster" --release-label emr-7.2.0 --
applications Name=Spark \
--ec2-attributes KeyName=myKey --instance-type m5.xlarge --instance-count 3 \
--steps Type=Spark,Name="Spark Program",ActionOnFailure=CONTINUE,Args=[--
class,org.apache.spark.examples.SparkPi,/usr/lib/spark/examples/jars/spark-
examples.jar,10] --use-default-roles
```

   As an alternative, you can use `command-runner.jar` as shown in the following example.

```
aws emr create-cluster --name "Add Spark Step Cluster" --release-label emr-7.2.0 \
--applications Name=Spark --ec2-attributes KeyName=myKey --instance-type m5.xlarge
 --instance-count 3 \
--steps Type=CUSTOM_JAR,Name="Spark Program",Jar="command-
runner.jar",ActionOnFailure=CONTINUE,Args=[spark-example,SparkPi,10] --use-default-
roles
```

   Note

   Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

2. Alternatively, add steps to a cluster already running. Use `add-steps`.

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF --steps
 Type=Spark,Name="Spark Program",ActionOnFailure=CONTINUE,Args=[--
class,org.apache.spark.examples.SparkPi,/usr/lib/spark/examples/jars/spark-
examples.jar,10]
```

As an alternative, you can use `command-runner.jar` as shown in the following example.

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF --steps Type=CUSTOM_JAR,Name="Spark
 Program",Jar="command-runner.jar",ActionOnFailure=CONTINUE,Args=[spark-
example,SparkPi,10]
```

To submit work to Spark using the SDK for Java

1. The following example shows how to add a step to a cluster with Spark using Java.

```
AWSCredentials credentials = new BasicAWSCredentials(accessKey, secretKey);
AmazonElasticMapReduce emr = new AmazonElasticMapReduceClient(credentials);

StepFactory stepFactory = new StepFactory();
AmazonElasticMapReduceClient emr = new AmazonElasticMapReduceClient(credentials);
AddJobFlowStepsRequest req = new AddJobFlowStepsRequest();
req.withJobFlowId("j-1K48XXXXXXHCB");

List<StepConfig> stepConfigs = new ArrayList<StepConfig>();

HadoopJarStepConfig sparkStepConf = new HadoopJarStepConfig()
   .withJar("command-runner.jar")
   .withArgs("spark-submit","--executor-memory","1g","--
class","org.apache.spark.examples.SparkPi","/usr/lib/spark/examples/jars/spark-
examples.jar","10");

StepConfig sparkStep = new StepConfig()
   .withName("Spark Step")
   .withActionOnFailure("CONTINUE")
   .withHadoopJarStep(sparkStepConf);

stepConfigs.add(sparkStep);
req.withSteps(stepConfigs);
AddJobFlowStepsResult result = emr.addJobFlowSteps(req);
```

2. View the results of the step by examining the logs for the step. You can do this in the AWS Management Console if you have enabled logging by choosing **Steps**, selecting your step, and then, for **Log files**, choosing either `stdout` or `stderr`. To see the logs available, choose **View Logs**.

Overriding Spark default configuration settings

You may want to override Spark default configuration values on a per-application basis. You can do this when you submit applications using a step, which essentially passes options to `spark-submit`. For example, you may wish to change the memory allocated to an executor process by changing `spark.executor.memory`. You would supply the `--executor-memory` switch with an argument like the following:

```
spark-submit --executor-memory 1g --class org.apache.spark.examples.SparkPi /usr/lib/spark/examples/jars/spark-examples.jar 10
```

Similarly, you can tune `--executor-cores` and `--driver-memory`. In a step, you would provide the following arguments to the step:

```
--executor-memory 1g --class org.apache.spark.examples.SparkPi /usr/lib/spark/examples/jars/spark-examples.jar 10
```

You can also tune settings that may not have a built-in switch using the `--conf` option. For more information about other settings that are tunable, see the [Dynamically loading Spark properties](#) topic in the Apache Spark documentation.

View Spark application history

You can view Spark, YARN application, and Tez UI details using the **Application user interfaces** tab of a cluster's detail page in the console. Amazon EMR application user interfaces (UI) make it easier for you to troubleshoot and analyze active jobs and job history.

For more information, see View application history in the *Amazon EMR Management Guide*.

Access the Spark web UIs

You can view the Spark web UIs by following the procedures to create an SSH tunnel or create a proxy in the section called Connect to the cluster in the Amazon EMR Management Guide and then navigating to the YARN ResourceManager for your cluster. Choose the link under **Tracking UI** for your application. If your application is running, you see **ApplicationMaster**. This takes you to the application master's web UI at port 20888 wherever the driver is located. The driver may be located on the cluster's primary node if you run in YARN client mode. If you are running an application in YARN cluster mode, the driver is located in the ApplicationMaster for the application on the cluster. If your application has finished, you see **History**, which takes you to the Spark HistoryServer UI port number at 18080 of the EMR cluster's primary node. This is for applications that have already completed. You can also navigate to the Spark HistoryServer UI directly at http://*master-public-dns-name*:18080/.

With Amazon EMR release 5.25.0 and later, you can access Spark history server UI from the console without setting up a web proxy through an SSH connection. For more information, see [View persistent application user interfaces](#).

Using the Spark structured streaming Amazon Kinesis Data Streams connector

Amazon EMR releases 7.1.0 and higher include a spark structured streaming Amazon Kinesis Data Streams connector in the release image. With this connector, you can use Spark on Amazon EMR to process data that's stored in Amazon Kinesis Data Streams. The connector supports both consumer types of `GetRecords` (shared throughput) and `SubscribeToShard` (enhanced fan-out). This integration is based on the `spark-sql-kinesis-connector`. For details about how to get started using the connector, see the README.

The following example demonstrates how to use the connector to launch a Spark application with Amazon EMR

```
spark-submit my_kinesis_streaming_script.py
```

Using Amazon Redshift integration for Apache Spark with Amazon EMR

With Amazon EMR release 6.4.0 and later, every release image includes a connector between Apache Spark and Amazon Redshift. With this connector, you can use Spark on Amazon EMR to process data stored in Amazon Redshift. For Amazon EMR releases 6.4.0 through 6.8.0, the integration is based on the `spark-redshift` open-source connector. For Amazon EMR releases 6.9.0 and later, the Amazon Redshift integration for Apache Spark has been migrated from the community version to a native integration.

Topics

Launching a Spark application using the Amazon Redshift integration for Apache Spark

For Amazon EMR releases 6.4 through 6.9, you must use the `--jars` or `--packages` option to specify which of the following JAR files you want to use. The `--jars` option specifies dependencies stored locally, in HDFS, or using HTTP/S. To see other file locations supported by the `--jars` option, see Advanced Dependency Management in the Spark documentation. The `--packages` option specifies dependencies stored in the public Maven repo.

- `spark-redshift.jar`
- `spark-avro.jar`
- `RedshiftJDBC.jar`
- `minimal-json.jar`

Amazon EMR releases 6.10.0 and higher don't require the `minimal-json.jar` dependency, and automatically install the other dependencies to each cluster by default. The following examples show how to launch a Spark application with the Amazon Redshift integration for Apache Spark.

Amazon EMR 6.10.0 +

The following example shows how to launch a Spark application with the `spark-redshift` connector with Amazon EMR releases 6.10 and higher.

```
spark-submit my_script.py
```

Amazon EMR 6.4.0 - 6.9.x

To launch a Spark application with the `spark-redshift` connector on Amazon EMR releases 6.4 through 6.9, you must use the `--jars` or `--packages` option, as the following example shows. Note that the paths listed with the `--jars` option are the default paths for the JAR files.

```
spark-submit \
  --jars /usr/share/aws/redshift/jdbc/RedshiftJDBC.jar,/usr/share/aws/redshift/
spark-redshift/lib/spark-redshift.jar,/usr/share/aws/redshift/spark-redshift/lib/
spark-avro.jar,/usr/share/aws/redshift/spark-redshift/lib/minimal-json.jar \
  my_script.py
```

Authenticating with Amazon Redshift integration for Apache Spark

Using AWS Secrets Manager to retrieve credentials and connect to Amazon Redshift

The following code sample shows how you can use AWS Secrets Manager to retrieve credentials to connect to an Amazon Redshift cluster with the PySpark interface for Apache Spark in Python.

```
from pyspark.sql import SQLContext
import boto3

sc = # existing SparkContext
sql_context = SQLContext(sc)

secretsmanager_client = boto3.client('secretsmanager')
secret_manager_response = secretsmanager_client.get_secret_value(
    SecretId='string',
    VersionId='string',
    VersionStage='string'
)
username = # get username from secret_manager_response
password = # get password from secret_manager_response
url = "jdbc:redshift://redshifthost:5439/database?user=" + username + "&password=" +
 password

# Read data from a table
df = sql_context.read \
    .format("io.github.spark_redshift_community.spark.redshift") \
    .option("url", url) \
```

```
    .option("dbtable", "my_table") \
    .option("tempdir", "s3://path/for/temp/data") \
    .load()
```

Using IAM to retrieve credentials and connect to Amazon Redshift

You can use the Amazon Redshift-provided JDBC version 2 driver to connect to Amazon Redshift with the Spark connector. To use AWS Identity and Access Management (IAM), configure your JDBC URL to use IAM authentication. To connect to a Redshift cluster from Amazon EMR, you must give your IAM role permission to retrieve temporary IAM credentials. Assign the following permissions to your IAM role so that it can retrieve credentials and run Amazon S3 operations.

- Redshift:GetClusterCredentials (for provisioned Amazon Redshift clusters)
- Redshift:DescribeClusters (for provisioned Amazon Redshift clusters)
- Redshift:GetWorkgroup (for Amazon Redshift Serverless workgroups)
- Redshift:GetCredentials (for Amazon Redshift Serverless workgroups)
- s3:GetBucket
- s3:GetBucketLocation
- s3:GetObject
- s3:PutObject
- s3:GetBucketLifecycleConfiguration

For more information about `GetClusterCredentials`, see Resource policies for `GetClusterCredentials`.

You also must make sure that Amazon Redshift can assume the IAM role during `COPY` and `UNLOAD` operations.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": "redshift.amazonaws.com"
            },
            "Action": "sts:AssumeRole"
        }
```

```
    ]
}
```

The following example uses IAM authentication between Spark and Amazon Redshift:

```
from pyspark.sql import SQLContext
import boto3

sc = # existing SparkContext
sql_context = SQLContext(sc)

url = "jdbc:redshift:iam//redshift-host:redshift-port/db-name"
iam_role_arn = "arn:aws:iam::account-id:role/role-name"

# Read data from a table
df = sql_context.read \
    .format("io.github.spark_redshift_community.spark.redshift") \
    .option("url", url) \
    .option("aws_iam_role", iam_role_arn) \
    .option("dbtable", "my_table") \
    .option("tempdir", "s3a://path/for/temp/data") \
    .mode("error") \
    .load()
```

Reading and writing from and to Amazon Redshift

The following code examples use PySpark to read and write sample data from and to an Amazon Redshift database with data source API and using SparkSQL.

Data source API

Use PySpark to read and write sample data from and to an Amazon Redshift database with data source API.

```
import boto3
from pyspark.sql import SQLContext

sc = # existing SparkContext
sql_context = SQLContext(sc)

url = "jdbc:redshift:iam://redshifthost:5439/database"
aws_iam_role_arn = "arn:aws:iam::accountID:role/roleName"
```

```
df = sql_context.read \
    .format("io.github.spark_redshift_community.spark.redshift") \
    .option("url", url) \
    .option("dbtable", "tableName") \
    .option("tempdir", "s3://path/for/temp/data") \
    .option("aws_iam_role", "aws_iam_role_arn") \
    .load()

df.write \
    .format("io.github.spark_redshift_community.spark.redshift") \
    .option("url", url) \
    .option("dbtable", "tableName_copy") \
    .option("tempdir", "s3://path/for/temp/data") \
    .option("aws_iam_role", "aws_iam_role_arn") \
    .mode("error") \
    .save()
```

SparkSQL

Use PySpark to read and write sample data from and to an Amazon Redshift database with SparkSQL.

```
import boto3
import json
import sys
import os
from pyspark.sql import SparkSession

spark = SparkSession \
    .builder \
    .enableHiveSupport() \
    .getOrCreate()

url = "jdbc:redshift:iam://redshifthost:5439/database"
aws_iam_role_arn = "arn:aws:iam::accountID:role/roleName"

bucket = "s3://path/for/temp/data"
tableName = "tableName" # Redshift table name

s = f"""CREATE TABLE IF NOT EXISTS {tableName} (country string, data string)
    USING io.github.spark_redshift_community.spark.redshift
```

```
    OPTIONS (dbtable '{tableName}', tempdir '{bucket}', url '{url}', aws_iam_role
 '{aws_iam_role_arn}' ); """

spark.sql(s)

columns = ["country" ,"data"]
data = [("test-country","test-data")]
df = spark.sparkContext.parallelize(data).toDF(columns)

# Insert data into table
df.write.insertInto(tableName, overwrite=False)
df = spark.sql(f"SELECT * FROM {tableName}")
df.show()
```

Considerations and limitations when using the Spark connector

- We recommend that you turn on SSL for the JDBC connection from Spark on Amazon EMR to Amazon Redshift.
- We recommend that you manage the credentials for the Amazon Redshift cluster in AWS Secrets Manager as a best practice. See Using AWS Secrets Manager to retrieve credentials for connecting to Amazon Redshift for an example.
- We recommend that you pass an IAM role with the parameter `aws_iam_role` for the Amazon Redshift authentication parameter.
- The `tempdir` URI points to an Amazon S3 location. This temp directory isn't cleaned up automatically and therefore could add additional cost.
- Consider the following recommendations for Amazon Redshift:
  - We recommend that you block public access to the Amazon Redshift cluster.
  - We recommend that you turn on Amazon Redshift audit logging.
  - We recommend that you turn on Amazon Redshift at-rest encryption.
- Consider the following recommendations for Amazon S3:
  - We recommend that you block public access to Amazon S3 buckets.
  - We recommend that you use Amazon S3 server-side encryption to encrypt the Amazon S3 buckets used.
  - We recommend that you use Amazon S3 lifecycle policies to define the retention rules for the Amazon S3 bucket.

- Amazon EMR always verifies code imported from open-source into the image. For security, we don't support the following authentication methods from Spark to Amazon S3:
  - Setting AWS access keys in the `hadoop-env` configuration classification
  - Encoding AWS access keys in the `tempdir` URI

For more information on using the connector and its supported parameters, see the following resources:

- Amazon Redshift integration for Apache Spark in the *Amazon Redshift Management Guide*
- The `spark-redshift` community repository on Github

Spark release history

The following table lists the version of Spark included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in Amazon EMR 7.x release versions, Amazon EMR 6.x release versions, or Amazon EMR 5.x release versions.

> **Important**
>
> Apache Spark version 2.3.1, available beginning with Amazon EMR release 5.16.0, addresses CVE-2018-8024 and CVE-2018-1334. We recommend that you migrate earlier versions of Spark to Spark version 2.3.1 or later.

Spark version information

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-7.2.0	3.5.1	delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hd fs-datanode, hadoop-hdfs-library, hadoop-hdfs-nameno de, hadoop-httpfs-server, hadoop-kms-server, hadoop-

Amazon EMR Release label	Spark Version	Components installed with Spark
		yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.36.2	2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-7.1.0	3.5.0	delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-7.0.0	3.5.0	delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-6.15.0	3.4.1	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.14.0	3.4.1	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-6.13.0	3.4.1	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.12.0	3.4.0	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-6.11.1	3.3.2	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.11.0	3.3.2	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-6.10.1	3.3.1	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.10.0	3.3.1	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-6.9.1	3.3.0	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.9.0	3.3.0	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-6.8.1	3.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.8.0	3.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-6.7.0	3.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.36.1	2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.36.0	2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.6.0	3.2.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.35.0	2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.5.0	3.1.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-6.4.0	3.1.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.3.1	3.1.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-6.3.0	3.1.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.2.1	3.0.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-6.2.0	3.0.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.1.1	3.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-6.1.0	3.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.0.1	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-6.0.0	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.34.0	2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.33.1	2.4.7	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.33.0	2.4.7	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.32.1	2.4.7	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.32.0	2.4.7	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.31.1	2.4.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.31.0	2.4.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.30.2	2.4.5	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.30.1	2.4.5	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.30.0	2.4.5	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-notebook-env, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.29.0	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.28.1	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.28.0	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.27.1	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.27.0	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.26.0	2.4.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.25.0	2.4.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.24.1	2.4.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.24.0	2.4.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.23.1	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.23.0	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.22.0	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.21.2	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.21.1	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.21.0	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.20.1	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.20.0	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.19.1	2.3.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.19.0	2.3.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.18.1	2.3.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.18.0	2.3.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.17.2	2.3.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.17.1	2.3.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.17.0	2.3.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.16.1	2.3.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.16.0	2.3.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.15.1	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.15.0	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.14.2	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.14.1	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.14.0	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.13.1	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.13.0	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.12.3	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.12.2	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.12.1	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.12.0	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.11.4	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.11.3	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.11.2	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.11.1	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.11.0	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.10.1	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.10.0	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.9.1	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.9.0	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.8.3	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.8.2	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.8.1	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.8.0	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hd fs-datanode, hadoop-hdfs-library, hadoop-hdfs-nameno de, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server , spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.7.1	2.1.1	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hd fs-datanode, hadoop-hdfs-library, hadoop-hdfs-nameno de, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server , spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.7.0	2.1.1	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.6.1	2.1.1	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.6.0	2.1.1	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.5.4	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.5.3	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.5.2	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.5.1	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.5.0	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.4.1	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.4.0	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.3.2	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.3.1	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.3.0	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.2.3	2.0.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.2.2	2.0.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.2.1	2.0.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.2.0	2.0.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.1.1	2.0.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.1.0	2.0.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.0.3	2.0.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.0.2	2.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.0.1	2.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-5.0.0	2.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.9.6	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-4.9.5	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.9.4	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-4.9.3	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.9.2	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-4.9.1	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.8.5	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-4.8.4	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.8.3	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-4.8.2	1.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.8.1	1.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-4.8.0	1.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.7.4	1.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-4.7.3	1.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.7.2	1.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-4.7.1	1.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.7.0	1.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.6.1	1.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-4.6.0	1.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.5.0	1.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.4.0	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-4.3.0	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.2.0	1.5.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.1.0	1.5.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR Release label	Spark Version	Components installed with Spark
emr-4.0.0	1.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Apache Sqoop

Apache Sqoop is a tool for transferring data between Amazon S3, Hadoop, HDFS, and RDBMS databases. For more information, see the Apache Sqoop website. Sqoop is included in Amazon EMR releases 5.0.0 and later. Earlier releases include Sqoop as a sandbox application. For more information, see Amazon EMR 4.x release versions.

Topics

Sqoop version information

Sqoop version for 7.2.0

The following table lists the version of Sqoop included in the latest release of the Amazon EMR 7.x series, along with the components that Amazon EMR installs with Sqoop.

For the version of components installed with Sqoop in this release, see Release 7.2.0 Component Versions.

Sqoop version information for emr-7.2.0

Amazon EMR Release Label	Sqoop Version	Components Installed With Sqoop
emr-7.2.0	Sqoop 1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-

Amazon EMR Release Label	Sqoop Version	Components Installed With Sqoop
		timeline-server, mariadb-server, sqoop-client

Sqoop version for 6.15.0

The following table lists the version of Sqoop included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with Sqoop.

For the version of components installed with Sqoop in this release, see Release 6.15.0 Component Versions.

Sqoop version information for emr-6.15.0

Amazon EMR Release Label	Sqoop Version	Components Installed With Sqoop
emr-6.15.0	Sqoop 1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Sqoop version for 5.36.2

The following table lists the version of Sqoop included in the latest release of the Amazon EMR 5.x series, along with the components that Amazon EMR installs with Sqoop.

For the version of components installed with Sqoop in this release, see Release 5.36.2 Component Versions.

Sqoop version information for emr-5.36.2

Amazon EMR Release Label	Sqoop Version	Components Installed With Sqoop
emr-5.36.2	Sqoop 1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Considerations with Sqoop on Amazon EMR

Consider the following items when you run Sqoop on Amazon EMR.

Using Sqoop with HCatalog integration

Sqoop on Amazon EMR supports Sqoop-HCatalog integration. When you use Sqoop to write output to an HCatalog table in Amazon S3, disable Amazon EMR direct write by setting the `mapred.output.direct.NativeS3FileSystem` and `mapred.output.direct.EmrFileSystem` properties to `false`. For more information, see Using HCatalog. You can use the Hadoop `-D mapred.output.direct.NativeS3FileSystem=false` and `-D mapred.output.direct.EmrFileSystem=false` commands. If you don't disable direct write, no error occurs, but the table is created in Amazon S3 and no data is written.

Sqoop JDBC and database support

By default, Sqoop has a MariaDB and PostgreSQL driver installed. The PostgreSQL driver installed for Sqoop only works for PostgreSQL 8.4. To install an alternate set of JDBC connectors for Sqoop, connect to the cluster master node and install them in `/usr/lib/sqoop/lib`. The following are links for various JDBC connectors:

- MariaDB: About MariaDB Connector/J.
- PostgreSQL: PostgreSQL JDBC driver.
- SQLServer: Download Microsoft JDBC driver for SQL Server.
- MySQL: Download Connector/J
- Oracle: Get Oracle JDBC drivers and UCP from the Oracle Maven repository

The supported databases for Sqoop are listed at the following url, http://sqoop.apache.org/docs/*version*/SqoopUserGuide.html#_supported_databases, where *version* is the version of Sqoop you are using, for example 1.4.6. If the JDBC connect string does not match those in this list, you must specify a driver.

For example, you can export to an Amazon Redshift database table with the following command (for JDBC 4.1):

```
sqoop export --connect jdbc:redshift://$MYREDSHIFTHOST:5439/mydb --table mysqoopexport
 --export-dir s3://mybucket/myinputfiles/ --driver com.amazon.redshift.jdbc41.Driver --
username master --password Mymasterpass1
```

You can use both the MariaDB and MySQL connection strings but if you specify the MariaDB connection string, you need to specify the driver:

```
sqoop export --connect jdbc:mariadb://$HOSTNAME:3306/mydb --table mysqoopexport
 --export-dir s3://mybucket/myinputfiles/ --driver org.mariadb.jdbc.Driver --
username master --password Mymasterpass1
```

If you are using Secure Socket Layer encryption to access your database, you need to use a JDBC URI like in the following Sqoop export example:

```
sqoop export --connect jdbc:mariadb://$HOSTNAME:3306/mydb?
verifyServerCertificate=false&useSSL=true&requireSSL=true --table mysqoopexport
 --export-dir s3://mybucket/myinputfiles/ --driver org.mariadb.jdbc.Driver --
username master --password Mymasterpass1
```

For more information about SSL encryption in RDS, see Using SSL to encrypt a connection to a DB instance in the Amazon RDS User Guide.

For more information, see the Apache Sqoop documentation.

Securing your password

There are several methods that you might choose from to securely pass your password:

Java KeyStore

The preferred method encrypts the password with a Java KeyStore (JKS), eliminating the need to store the password in a readable format.

1. Create a password alias. On prompt, enter the password that you use to access the database.

```
hadoop credential create mydb.password.alias -provider jceks://hdfs/user/root/
mysql.password.jceks
```

2. Use the password alias to launch the Sqoop job:

```
sqoop export -Dhadoop.security.credential.provider.path=jceks://hdfs/user/
root/mysql.password.jceks --connect jdbc:mariadb://$HOSTNAME:3306/mydb
 --table mysqoopexport --export-dir s3://mybucket/myinputfiles/ --driver
 org.mariadb.jdbc.Driver --username master --password-alias mydb.password.alias
```

--password-file

You can use the `--password-file` command to pass the password through a file as shown in the following example:

1. Create a new file that contains the password:

```
echo -n 'Mymasterpass1' > /home/hadoop/mysql-pass.password
```

2. Use the file to launch the Sqoop job:

```
sqoop export --connect jdbc:mariadb://$HOSTNAME:3306/mydb --table mysqoopexport
 --export-dir s3://mybucket/myinputfiles/ --driver org.mariadb.jdbc.Driver --
username master --password-file /home/hadoop/mysql-pass.password
```

-P

You can use the -P command to enter the password through a prompt as shown in the following example:

```
sqoop export --connect jdbc:mariadb://$HOSTNAME:3306/mydb --table mysqoopexport --
export-dir s3://mybucket/myinputfiles/ --driver org.mariadb.jdbc.Driver --username
 master -P
```

Sqoop release history

The following table lists the version of Sqoop included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in Amazon EMR 7.x release versions, Amazon EMR 6.x release versions, or Amazon EMR 5.x release versions.

Sqoop version information

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-7.2.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.36.2	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
		httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-7.1.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-7.0.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-6.15.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.14.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-6.13.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.12.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-6.11.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.11.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-6.10.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.10.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-6.9.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.9.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-6.8.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.8.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-6.7.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.36.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.36.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.6.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.35.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.5.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-6.4.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.3.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-6.3.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.2.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-6.2.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.1.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-6.1.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.34.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.33.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.33.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.32.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.32.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.31.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.31.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.30.2	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.30.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.30.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.29.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.28.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.28.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.27.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.27.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.26.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.25.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.24.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.24.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.23.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.23.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.22.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.21.2	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.21.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.21.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.20.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.20.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.19.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.19.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.18.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.18.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.17.2	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.17.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.17.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.16.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.16.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.15.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.15.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.14.2	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.14.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.14.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.13.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.13.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.12.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.12.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.12.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.12.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.11.4	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.11.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.11.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.11.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.11.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.10.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.10.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.9.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.9.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.8.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.8.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.8.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.8.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.7.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.7.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.6.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.6.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.5.4	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.5.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.5.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.5.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.5.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.4.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.4.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.3.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.3.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.3.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.2.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.2.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.2.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.2.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.1.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.1.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.0.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon EMR Release label	Sqoop Version	Components installed with Sqoop
emr-5.0.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.0.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.0.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

TensorFlow

TensorFlow is an open-source symbolic math library for machine intelligence and deep learning applications. For more information, see the TensorFlow website. TensorFlow is available with Amazon EMR release version 5.17.0 and later.

The following table lists the version of TensorFlow included in the latest release of the Amazon EMR 7.x series, along with the components that Amazon EMR installs with TensorFlow.

For the version of components installed with TensorFlow in this release, see Release 7.2.0 Component Versions.

TensorFlow version information for emr-7.2.0

Amazon EMR Release Label	TensorFlow Version	Components Installed With TensorFlow
emr-7.2.0	TensorFlow 2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

The following table lists the version of TensorFlow included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with TensorFlow.

For the version of components installed with TensorFlow in this release, see Release 6.15.0 Component Versions.

TensorFlow version information for emr-6.15.0

Amazon EMR Release Label	TensorFlow Version	Components Installed With TensorFlow
emr-6.15.0	TensorFlow 2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

The following table lists the version of TensorFlow included in the latest release of the Amazon EMR 5.x series, along with the components that Amazon EMR installs with TensorFlow.

For the version of components installed with TensorFlow in this release, see Release 5.36.2 Component Versions.

TensorFlow version information for emr-5.36.2

Amazon EMR Release Label	TensorFlow Version	Components Installed With TensorFlow
emr-5.36.2	TensorFlow 2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

TensorFlow builds by Amazon EC2 instance type

Amazon EMR uses different builds of the TensorFlow library depending on the instance types that you choose for your cluster. The following table lists builds by instance type.

EC2 instance types	TensorFlow build
M5 and C5	Tensorflow 1.9.0 with Intel MKL optimization
P2	Tensorflow 1.9.0 with CUDA 9.2, cuDNN 7.1
P3	Tensorflow 1.9.0 with CUDA 9.2, cuDNN 7.1, NCCL 2.2.13 Nvidia NCCL is available only on P3 instances. **End User License Agreement (EULA)**: By using Nvidia components on Amazon EMR, you agree to the terms and conditions outlined in the product EULA.
All others	Tensorflow 1.9.0

Security

In addition to following the guidance in Using TensorFlow securely we recommend that you launch your cluster in a private subnet to help you limit access to trusted sources. For more information, see Amazon VPC options in the *Amazon EMR Management Guide*.

Using TensorBoard

TensorBoard is a suite of visualization tools for TensorFlow programs. For more information, see TensorBoard: Visualized learning on the Tensorflow website.

To use TensorBoard with Amazon EMR, you must start TensorBoard on the cluster master node.

To use tensorboard with Tensorflow on Amazon EMR

1. Connect to the master node of the cluster using SSH. For more information, see Connect to the master node using SSH in the *Amazon EMR Management Guide*.
2. Type the following command to start Tensorboard on the master node. Replace */my/log/directory* with a directory on the master node where you have generated and stored summary data using a summary writer.

   Amazon EMR 5.19.0 and later

   ```
   python3 -m tensorboard.main --logdir=/home/hadoop/tensor --bind_all
   ```

   Amazon EMR 5.18.1 and earlier

   ```
   python3 -m tensorboard.main --logdir=/my/log/dir
   ```

   By default, the master node hosts TensorBoard using port 6006 and the master public DNS name. After you start TensorBoard, the command line output presents the URL that can be used to connect to TensorBoard, as shown in the following example:

   ```
   TensorBoard 1.9.0 at http://master-public-dns-name:6006 (Press CTRL+C to quit)
   ```
3. Set up access to web interfaces on the master node from trusted clients. For more information, see View web interfaces hosted on Amazon EMR clusters in the *Amazon EMR Management Guide*.
4. Open TensorBoard at `http://master-public-dns-name:6006`.

TensorFlow release history

The following table lists the version of TensorFlow included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in Amazon EMR 7.x release versions, Amazon EMR 6.x release versions, or Amazon EMR 5.x release versions.

TensorFlow version information

Amazon EMR Release label	TensorFlow Version	Components installed with TensorFlow
emr-7.2.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.36.2	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-7.1.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR Release label	TensorFlow Version	Components installed with TensorFlow
emr-7.0.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.15.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.14.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR Release label	TensorFlow Version	Components installed with TensorFlow
emr-6.13.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.12.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.11.1	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR Release label	TensorFlow Version	Components installed with TensorFlow
emr-6.11.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.10.1	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.10.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR Release label	TensorFlow Version	Components installed with TensorFlow
emr-6.9.1	2.10.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.9.0	2.10.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.8.1	2.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR Release label	TensorFlow Version	Components installed with TensorFlow
emr-6.8.0	2.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.7.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.36.1	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR Release label	TensorFlow Version	Components installed with TensorFlow
emr-5.36.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.6.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.35.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR Release label	TensorFlow Version	Components installed with TensorFlow
emr-6.5.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.4.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.3.1	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR Release label	TensorFlow Version	Components installed with TensorFlow
emr-6.3.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.2.1	2.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.2.0	2.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR Release label	TensorFlow Version	Components installed with TensorFlow
emr-6.1.1	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.1.0	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.0.1	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR Release label	TensorFlow Version	Components installed with TensorFlow
emr-6.0.0	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.34.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.33.1	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR Release label	TensorFlow Version	Components installed with TensorFlow
emr-5.33.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.32.1	2.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.32.0	2.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR Release label	TensorFlow Version	Components installed with TensorFlow
emr-5.31.1	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.31.0	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.30.2	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR Release label	TensorFlow Version	Components installed with TensorFlow
emr-5.30.1	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.30.0	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.29.0	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR Release label	TensorFlow Version	Components installed with TensorFlow
emr-5.28.1	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.28.0	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.27.1	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR Release label	TensorFlow Version	Components installed with TensorFlow
emr-5.27.0	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.26.0	1.13.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.25.0	1.13.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR Release label	TensorFlow Version	Components installed with TensorFlow
emr-5.24.1	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.24.0	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.23.1	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR Release label	TensorFlow Version	Components installed with TensorFlow
emr-5.23.0	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.22.0	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.21.2	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR Release label	TensorFlow Version	Components installed with TensorFlow
emr-5.21.1	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.21.0	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.20.1	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR Release label	TensorFlow Version	Components installed with TensorFlow
emr-5.20.0	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.19.1	1.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.19.0	1.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR Release label	TensorFlow Version	Components installed with TensorFlow
emr-5.18.1	1.9.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.18.0	1.9.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.17.2	1.9.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR Release label	TensorFlow Version	Components installed with TensorFlow
emr-5.17.1	1.9.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.17.0	1.9.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Apache Tez

Apache Tez is a framework that creates a complex directed acyclic graph (DAG) of tasks for processing data. You can use it as an alternative to Hadoop MapReduce for some use cases. For example, you can run Pig and Hive workflows with Hadoop MapReduce or you can use Tez as an execution engine. For more information, see https://tez.apache.org/. Amazon EMR releases 4.7.0 and higher include Tez.

The following table lists the version of Tez included in the latest release of the Amazon EMR 7.x series, along with the components that Amazon EMR installs with Tez.

For the version of components installed with Tez in this release, see Release 7.2.0 Component Versions.

Tez version information for emr-7.2.0

Amazon EMR Release Label	Tez Version	Components Installed With Tez
emr-7.2.0	Tez 0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker

The following table lists the version of Tez included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with Tez.

For the version of components installed with Tez in this release, see Release 6.15.0 Component Versions.

Tez version information for emr-6.15.0

Amazon EMR Release Label	Tez Version	Components Installed With Tez
emr-6.15.0	Tez 0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker

The following table lists the version of Tez included in the latest release of the Amazon EMR 5.x series, along with the components that Amazon EMR installs with Tez.

For the version of components installed with Tez in this release, see Release 5.36.2 Component Versions.

Tez version information for emr-5.36.2

Amazon EMR Release Label	Tez Version	Components Installed With Tez
emr-5.36.2	Tez 0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Topics

Creating a cluster with Tez

To install Tez, choose Apache Tez as an application when you create your cluster.

To create a cluster with Tez installed using the console

1. Navigate to the new Amazon EMR console and select **Switch to the old console** from the side navigation. For more information on what to expect when you switch to the old console, see Using the old console.
2. Choose **Create cluster**, **Go to advanced options**.
3. Under **Software Configuration**, select a **Release** of **emr-4.7.0** or higher.
4. Select **Tez** along with other applications you want Amazon EMR to install.
5. Select other options as necessary and then choose **Create cluster**.

To create a cluster with Tez using the AWS CLI

- Use the `create-cluster` command along with the `-- applications` option to specify **Tez**. The following example creates a cluster with Tez installed.

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --name "Cluster with Tez" --release-label emr-7.2.0 \
--applications Name=Tez --ec2-attributes KeyName=myKey \
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

Configuring Tez

You can customize Tez by setting values using the `tez-site` configuration classification, which configures settings in the `tez-site.xml` configuration file. For more information, see [TezConfiguration](#) in the Apache Tez documentation. To change Hive or Pig to use the Tez execution engine, use the `hive-site` and `pig-properties` configuration classifications as appropriate. Examples are shown below.

Example configuration

Example Example: Customizing the Tez root logging level and setting Tez as the execution engine for Hive and Pig

The example `create-cluster` command shown below creates a cluster with Tez, Hive, and Pig installed. The command references a file stored in Amazon S3, `myConfig.json`, which specifies properties for the `tez-site` classification that sets `tez.am.log.level` to DEBUG, and sets the execution engine to Tez for Hive and Pig using the `hive-site` and `pig-properties` configuration classifications.

> **Note**
>
> Linux line continuation characters (\) are included for readability. They can be removed or used in Linux commands. For Windows, remove them or replace with a caret (^).

```
aws emr create-cluster --release-label emr-7.2.0 \
--applications Name=Tez Name=Hive Name=Pig --ec2-attributes KeyName=myKey \
--instance-type m5.xlarge --instance-count 3 \
--configurations https://s3.amazonaws.com/mybucket/myfolder/myConfig.json --use-
default-roles
```

Example contents of `myConfig.json` are shown below.

```
[
    {
      "Classification": "tez-site",
      "Properties": {
        "tez.am.log.level": "DEBUG"
```

```
      }
    },
    {
      "Classification": "hive-site",
      "Properties": {
        "hive.execution.engine": "tez"
      }
    },
    {
      "Classification": "pig-properties",
      "Properties": {
        "exectype": "tez"
      }
    }
  ]
```

Note

With Amazon EMR version 5.21.0 and later, you can override cluster configurations and specify additional configuration classifications for each instance group in a running cluster. You do this by using the Amazon EMR console, the AWS Command Line Interface (AWS CLI), or the AWS SDK. For more information, see Supplying a Configuration for an Instance Group in a Running Cluster.

Tez asynchronous split opening

When there is a large number of small files in the table path, and a query attempts to read them all, each small file that corresponds to each individual split gets combined under one Tez *grouped split*. A single mapper then processes the single Tez grouped split. Since the execution is synchronous, each individual split under the grouped split gets processed one by one. This requires `RecordReader` objects to synchronously process the splits.

Name	Classification	Description
`tez.grouping.split.init.threads`	`tez-site`	Specifies the number of daemon threads that Tez uses to pre-initiate the `RecordReaders` and open

Name	Classification	Description
		splits. For ACID tables, the maximum supported value of `tez.grouping.split.init.threads` is 1.
`tez.grouping.split.init.recordreaders`	`tez-site`	Specifies the number of `RecordReaders` to keep pre-initialised by the daemon threads. This can help when Tez grouped split contains a large number of `InputSplits`. Initialization of `RecordReaders` to process those input splits can be done asynchronously with daemon threads instead of sequential processing.

Benchmarking for Tez asynchronous split opening

We used the following environments and configurations for benchmarking the Tez asynchronous split opening capability:

- **Benchmark environment** – Amazon EMR cluster with 1 primary node that uses m5.16xlarge, and 16 core nodes that use m5.16xlarge.
- **Benchmark configurations** – To simulate the scenario for benchmarking where a large number of input splits are in a single Tez grouped split, `tez.grouping.split-count` is set to 1.
- **Table used for benchmarking** – The table contains 200 partitions, with each partition containing a single file. The benchmark is done for when that table contains CSV files, and when that table contains parquet files. Hive query for benchmarking: `SELECT COUNT(*)` from the table ten times, and take the average runtime.
- **Configurations to enable Tez async split opening** – As follows:
  - `tez.grouping.split.init.threads` = 4
  - `tez.grouping.split.init.recordreaders` = 10

Dataset	Feature disabled (baseline)	Feature enabled	Improvement
CSV dataset	90.26 seconds	79.20 seconds	12.25%
Parquet dataset	54.67 seconds	42.23 seconds	22.75%

Tez web UI

Tez has its own web user interface. To view the web UI, see the following URL.

```
http://masterDNS:8080/tez-ui
```

To enable the Hive Queries tab on the Tez web UI, set the following configuration.

```
[
  {
    "Classification": "hive-site",
    "Properties": {
      "hive.exec.pre.hooks": "org.apache.hadoop.hive.ql.hooks.ATSHook",
      "hive.exec.post.hooks": "org.apache.hadoop.hive.ql.hooks.ATSHook",
      "hive.exec.failure.hooks": "org.apache.hadoop.hive.ql.hooks.ATSHook"
    }
  }
]
```

You can also view Tez, Spark, and YARN application UI details using links on the **Application user interfaces** tab of a cluster's detail page in the console. Amazon EMR application user interfaces (UI) are hosted off-cluster and are available after the cluster has terminated. They don't require you to set up a SSH connection or web proxy, making it easier for you to troubleshoot and analyze active jobs and job history.

For more information, see View application history in the *Amazon EMR Management Guide*.

Timeline Server

The YARN Timeline Server is configured to run when Tez is installed. To view jobs submitted through Tez or MapReduce execution engines using the Timeline Server, view the web UI using the

URL `http://master-public-DNS:8188`. For more information, see View web interfaces hosted on Amazon EMR clusters in the *Amazon EMR Management Guide*.

Tez release history

The following table lists the version of Tez included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in Amazon EMR 7.x release versions, Amazon EMR 6.x release versions, or Amazon EMR 5.x release versions.

Tez version information

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-7.2.0	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-5.36.2	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-7.1.0	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-7.0.0	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-6.15.0	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-6.14.0	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-6.13.0	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-6.12.0	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-6.11.1	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-6.11.0	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-6.10.1	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-6.10.0	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-6.9.1	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.9.0	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-6.8.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.8.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.7.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.36.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.36.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.6.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.35.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.5.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.4.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-6.3.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.3.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.2.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-6.2.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.1.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.1.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-6.0.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.0.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.34.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.33.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.33.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.32.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.32.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.31.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.31.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.30.2	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.30.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.30.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.29.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.28.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.28.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.27.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.27.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.26.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.25.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.24.1	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.24.0	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.23.1	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.23.0	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.22.0	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.21.2	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.21.1	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.21.0	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.20.1	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.20.0	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.19.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.19.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.18.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.18.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.17.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.17.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.17.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.16.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.16.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.15.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.15.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.14.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.14.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.14.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.13.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.13.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.12.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.12.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.12.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.12.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.11.4	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.11.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.11.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.11.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.11.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.10.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.10.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.9.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.9.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.8.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.8.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.8.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.8.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.7.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.7.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.6.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.6.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.5.4	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.5.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.5.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.5.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.5.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.4.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.4.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.3.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.3.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.3.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.2.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.2.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.2.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.2.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.1.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.1.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.0.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.0.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-5.0.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.0.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.9.6	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-4.9.5	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.9.4	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.9.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-4.9.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.9.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.8.5	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-4.8.4	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.8.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.8.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-4.8.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.8.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.7.4	0.8.3	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-4.7.3	0.8.3	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.7.2	0.8.3	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.7.1	0.8.3	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR Release label	Tez Version	Components installed with Tez
emr-4.7.0	0.8.3	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Tez release notes by version

Topics

Amazon EMR 7.1.0 - Tez release notes

Amazon EMR 7.1.0 - Tez changes

Type	Description
Bug Fix	TEZ-4394 – Netty4 ShuffleHandler: should use 1 boss thread (#256)
Upgrade	Upgrades Netty to 4.1.100.Final.
Upgrade	Upgrades Jetty to 9.4.53.v20231009.

Amazon EMR 6.15.0 - Tez release notes

Amazon EMR 6.15.0 - Tez changes

Type	Description
Feature	TEZ-4397: Open Tez Input splits asynchron ously
Upgrade	TEZ-4493: Upgrade Apache Hadoop to 3.3.6

Amazon EMR 6.15.0 - Tez features

- **Tez asynchronous split opening** – Amazon EMR 6.15.0 introduces configurations that you can specify to asynchronously open the input splits in a Tez *grouped split*. The feature was initiated by TEZ-4397, but had regressions in OSS Hive. Amazon EMR Hive fixed the regressions and additional bugs in Hive ACID table. This improvement results in faster performance of read queries when there are a large number of input splits in a single Tez grouped split. For more information, see Tez asynchronous split opening.

Amazon EMR 6.14.0 - Tez release notes

Amazon EMR 6.14.0 - Tez changes

Type	Description
Improvement	Upgrade TLS version in Tez to 1.2

Amazon EMR 6.13.0 - Tez release notes

Amazon EMR 6.13.0 - Tez changes

Type	Description
Bug Fix	REVERT TEZ-4295: Could not decompress data. Buffer length is too small.
Bug Fix	REVERT TEZ-4302: `NullPointerException` in `CodecUtils` with `GzipCodec` .
Bug Fix	REVERT TEZ-4234: Compressor can cause `IllegalArgumentException` in `Buffer.limit` where limit exceeds capacity.
Bug Fix	REVERT TEZ-4135: Improve memory allocation when executing in-memory reads.

Amazon EMR 6.12.0 - Tez release notes

Amazon EMR 6.12.0 - Tez changes

Type	Description
Improvement	Added Support For JDK 11 and JDK 17 Runtime

Type	Description
Bug Fix	TEZ-4492: Update Bowerrc to use bower.herokuapp mirror to avoid Bower Registry CERT_EXPIRE issue (BOWER-2608)
Upgrade	Upgraded Surefire to 3.0.0-M7

Amazon EMR 6.11.0 - Tez release notes

Amazon EMR 6.11.0 - Tez changes

Type	Description
Bug	Fixed invalid vertex state transition during vertex level cleanup of shuffle data
Bug	Fixed DAG or vertex level cleanup of shuffle data not working
Improvement	Enable tez.am.dag.cleanup.on.completion by default for clearing shuffle data of completed DAGs

Amazon EMR 6.10.0 - Tez release notes

Amazon EMR 6.10.0 - Tez changes

Type	Description
Feature	Enable `tez.runtime.transfer.data-via-events.enabled` by default
Backport	TEZ-4450: Fix shuffle data fetch failure when shuffle data is transferred via data movement events

Type	Description
Backport	TEZ-4460: Fix read timeout error in fetching shuffle data from Tez Shuffle Handler
Backport	TEZ-4455: Add LoggingHandler in ShuffleHandler pipeline for better debuggability
Bug	Fix Tez task getting stuck intermittently when preemption of task is enabled

Amazon EMR 6.9.0 - Tez release notes

Amazon EMR 6.9.0 - Tez changes

Type	Description
Upgrade	Tez is upgraded to 0.10.2. For more information, see the change log for Apache Tez 0.10.2.
Upgrade	Upgrade Hadoop to 3.3.3.
Bug	Disable `tez.runtime.transfer.data-via-events.enabled` by default due to TEZ-4450.

Amazon EMR 6.8.0 - Tez release notes

Amazon EMR 6.8.0 - Tez changes

Type	Description
Backport	TEZ-3363: Delete intermediate data at the vertex level for Shuffle Handler
Backport	TEZ-4129: Delete intermediate attempt data for failed attempts for Shuffle Handler

Type	Description
Backport	TEZ-4430: Fixed tez.task.launch.cmd-opts property not working

Amazon EMR 6.7.0 - Tez release notes

Amazon EMR 6.7.0 - Tez changes

Type	Description
Backport	TEZ-4403: Upgrade SLF4J version to 1.7.36
Backport	TEZ-4405: Replace log4j 1.x with reload4j
Backport	TEZ-4411: Tez Build Failure: FileSaver.js not found

Amazon EMR 6.6.0 - Tez release notes

Amazon EMR 6.6.0 - Tez changes

Type	Description
Backport	TEZ-3918: Fixed tez.task.log.level property not working.
Backport	TEZ-4353: Update commons-io to 2.8.0.
Backport	TEZ-4114: Remove direct jetty dependency from tez.
Backport	TEZ-4323: Jetty jars were removed from dist package with TEZ-4114.

Apache Zeppelin

Use Apache Zeppelin as a notebook for interactive data exploration. For more information about Zeppelin, see https://zeppelin.apache.org/. Zeppelin is included in Amazon EMR releases 5.0.0 and later. Earlier releases include Zeppelin as a sandbox application. For more information, see Amazon EMR 4.x release versions.

To access the Zeppelin web interface, set up an SSH tunnel to the master node and a proxy connection. For more information, see View web interfaces hosted on EMR clusters.

The following table lists the version of Zeppelin included in the latest release of the Amazon EMR 7.x series, along with the components that Amazon EMR installs with Zeppelin.

For the version of components installed with Zeppelin in this release, see Release 7.2.0 Component Versions.

Zeppelin version information for emr-7.2.0

Amazon EMR Release Label	Zeppelin Version	Components Installed With Zeppelin
emr-7.2.0	Zeppelin 0.10.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

The following table lists the version of Zeppelin included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with Zeppelin.

For the version of components installed with Zeppelin in this release, see Release 6.15.0 Component Versions.

Zeppelin version information for emr-6.15.0

Amazon EMR Release Label	Zeppelin Version	Components Installed With Zeppelin
emr-6.15.0	Zeppelin 0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

The following table lists the version of Zeppelin included in the latest release of the Amazon EMR 5.x series, along with the components that Amazon EMR installs with Zeppelin.

For the version of components installed with Zeppelin in this release, see Release 5.36.2 Component Versions.

Zeppelin version information for emr-5.36.2

Amazon EMR Release Label	Zeppelin Version	Components Installed With Zeppelin
emr-5.36.2	Zeppelin 0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-nameno

Amazon EMR Release Label	Zeppelin Version	Components Installed With Zeppelin
		de, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server , hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Topics

- Considerations when using Zeppelin on Amazon EMR
- Zeppelin release history

Considerations when using Zeppelin on Amazon EMR

- Connect to Zeppelin using the same SSH tunneling method to connect to other web servers on the master node. Zeppelin server is found at port 8890.
- Zeppelin on Amazon EMR release versions 5.0.0 and later supports Shiro authentication.
- Zeppelin on Amazon EMR release versions 5.8.0 and later supports using AWS Glue Data Catalog as the metastore for Spark SQL. For more information, see Using AWS Glue Data Catalog as the metastore for Spark SQL.
- Zeppelin does not use some of the settings defined in your cluster's `spark-defaults.conf` configuration file, even though it instructs YARN to allocate executors dynamically if you have set `spark.dynamicAllocation.enabled` to `true`. You must set executor settings, such as memory and cores, using the Zeppelin **Interpreter** tab, and then restart the interpreter for them to be used.
- Amazon EMR releases 6.10.0 and higher support Apache Zeppelin integration with Apache Flink. See Working with Flink jobs from Zeppelin in Amazon EMR for more information.
- Zeppelin on Amazon EMR does not support the SparkR interpreter.

Zeppelin release history

The following table lists the version of Zeppelin included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in Amazon EMR 7.x release versions, Amazon EMR 6.x release versions, or Amazon EMR 5.x release versions.

Zeppelin version information

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-7.2.0	0.10.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.36.2	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
		server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-7.1.0	0.10.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-7.0.0	0.10.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-6.15.0	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.14.0	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-6.13.0	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.12.0	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-6.11.1	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.11.0	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-6.10.1	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.10.0	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-6.9.1	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.9.0	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-6.8.1	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.8.0	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-6.7.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.36.1	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.36.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.6.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.35.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.5.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-6.4.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.3.1	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-6.3.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.2.1	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-6.2.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.1.1	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-6.1.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.0.1	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-6.0.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.34.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.33.1	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.33.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.32.1	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.32.0	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.31.1	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.31.0	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.30.2	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.30.1	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.30.0	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.29.0	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.28.1	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.28.0	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.27.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.27.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.26.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.25.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.24.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.24.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.23.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.23.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.22.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.21.2	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.21.1	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.21.0	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.20.1	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.20.0	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.19.1	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.19.0	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.18.1	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.18.0	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.17.2	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.17.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.17.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.16.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.16.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.15.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.15.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.14.2	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.14.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.14.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.13.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.13.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.12.3	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.12.2	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.12.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.12.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.11.4	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.11.3	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.11.2	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.11.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.11.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.10.1	0.7.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.10.0	0.7.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.9.1	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.9.0	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.8.3	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.8.2	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.8.1	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.8.0	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.7.1	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.7.0	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.6.1	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.6.0	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.5.4	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.5.3	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.5.2	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.5.1	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.5.0	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.4.1	0.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.4.0	0.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.3.2	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.3.1	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.3.0	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.2.3	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.2.2	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.2.1	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.2.0	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.1.1	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.1.0	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.0.3	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.0.2	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.0.1	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR Release label	Zeppelin Version	Components installed with Zeppelin
emr-5.0.0	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Apache ZooKeeper

Apache ZooKeeper is a centralized service for maintaining configuration information, naming, providing distributed synchronization, and providing group services. For more information about ZooKeeper, see http://zookeeper.apache.org/.

The following table lists the version of ZooKeeper included in the latest release of the Amazon EMR 7.x series, along with the components that Amazon EMR installs with ZooKeeper.

For the version of components installed with ZooKeeper in this release, see Release 7.2.0 Component Versions.

ZooKeeper version information for emr-7.2.0

Amazon EMR Release Label	ZooKeeper Version	Components Installed With ZooKeeper
emr-7.2.0	ZooKeeper 3.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

The following table lists the version of ZooKeeper included in the latest release of the Amazon EMR 6.x series, along with the components that Amazon EMR installs with ZooKeeper.

For the version of components installed with ZooKeeper in this release, see Release 6.15.0 Component Versions.

ZooKeeper version information for emr-6.15.0

Amazon EMR Release Label	ZooKeeper Version	Components Installed With ZooKeeper
emr-6.15.0	ZooKeeper 3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

The following table lists the version of ZooKeeper included in the latest release of the Amazon EMR 5.x series, along with the components that Amazon EMR installs with ZooKeeper.

For the version of components installed with ZooKeeper in this release, see Release 5.36.2 Component Versions.

ZooKeeper version information for emr-5.36.2

Amazon EMR Release Label	ZooKeeper Version	Components Installed With ZooKeeper
emr-5.36.2	ZooKeeper 3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server,

Amazon EMR Release Label	ZooKeeper Version	Components Installed With ZooKeeper
		zookeeper-client, zookeeper-server

Topics

- ZooKeeper release history

ZooKeeper release history

The following table lists the version of ZooKeeper included in each release version of Amazon EMR, along with the components installed with the application. For component versions in each release, see the Component Version section for your release in Amazon EMR 7.x release versions, Amazon EMR 6.x release versions, or Amazon EMR 5.x release versions.

ZooKeeper version information

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-7.2.0	3.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.36.2	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-nameno

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
		de, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-7.1.0	3.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-7.0.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-6.15.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.14.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-6.13.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.12.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-6.11.1	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.11.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-6.10.1	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.10.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-6.9.1	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.9.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-6.8.1	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.8.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-6.7.0	3.5.7	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.36.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.36.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.6.0	3.5.7	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.35.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.5.0	3.5.7	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-6.4.0	3.5.7	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.3.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-6.3.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.2.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-6.2.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.1.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-6.1.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.0.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-6.0.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.34.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.33.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.33.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.32.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.32.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.31.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.31.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.30.2	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.30.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.30.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.29.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.28.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.28.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.27.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.27.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.26.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.25.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.24.1	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.24.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.23.1	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.23.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.22.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.21.2	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.21.1	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.21.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.20.1	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.20.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.19.1	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.19.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.18.1	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.18.0	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.17.2	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.17.1	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.17.0	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.16.1	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.16.0	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.15.1	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.15.0	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.14.2	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.14.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.14.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.13.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.13.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.12.3	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.12.2	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.12.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.12.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.11.4	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.11.3	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.11.2	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.11.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.11.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.10.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.10.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.9.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.9.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.8.3	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.8.2	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.8.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.8.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.7.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.7.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.6.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.6.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.5.4	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.5.3	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.5.2	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.5.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.5.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.4.1	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.4.0	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.3.2	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.3.1	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.3.0	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.2.3	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.2.2	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.2.1	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.2.0	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.1.1	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.1.0	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.0.3	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Amazon EMR Release label	ZooKeeper Version	Components installed with ZooKeeper
emr-5.0.2	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.0.1	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.0.0	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Connectors and utilities

Amazon EMR provides several connectors and utilities to access other AWS services as data sources. You can usually access data in these services within a program. For example, you can specify an Kinesis stream in a Hive query, Pig script, or MapReduce application and then operate on that data.

Topics

Export, import, query, and join tables in DynamoDB using Amazon EMR

> **Note**
>
> The Amazon EMR-DynamoDB Connector is open-sourced on GitHub. For more information, see https://github.com/awslabs/emr-dynamodb-connector.

DynamoDB is a fully managed NoSQL database service that provides fast and predictable performance with seamless scalability. Developers can create a database table and grow its request traffic or storage without limit. DynamoDB automatically spreads the data and traffic for the table over a sufficient number of servers to handle the request capacity specified by the customer and the amount of data stored, while maintaining consistent, fast performance. Using Amazon EMR and Hive you can quickly and efficiently process large amounts of data, such as data stored in DynamoDB. For more information about DynamoDB, see Amazon DynamoDB Developer Guide.

Apache Hive is a software layer that you can use to query map reduce clusters using a simplified, SQL-like query language called HiveQL. It runs on top of the Hadoop architecture. For more information about Hive and HiveQL, go to the HiveQL language manual. For more information about Hive and Amazon EMR, see Apache Hive .

You can use Amazon EMR with a customized version of Hive that includes connectivity to DynamoDB to perform operations on data stored in DynamoDB:

- Loading DynamoDB data into the Hadoop Distributed File System (HDFS) and using it as input into an Amazon EMR cluster.
- Querying live DynamoDB data using SQL-like statements (HiveQL).
- Joining data stored in DynamoDB and exporting it or querying against the joined data.
- Exporting data stored in DynamoDB to Amazon S3.
- Importing data stored in Amazon S3 to DynamoDB.

> **Note**
>
> The Amazon EMR-DynamoDB Connector does not support clusters configured to use Kerberos authentication.

To perform each of the following tasks, you'll launch an Amazon EMR cluster, specify the location of the data in DynamoDB, and issue Hive commands to manipulate the data in DynamoDB.

There are several ways to launch an Amazon EMR cluster: you can use the Amazon EMR console, the command line interface (CLI), or you can program your cluster using an AWS SDK or the Amazon EMR API. You can also choose whether to run a Hive cluster interactively or from a script. In this section, we will show you how to launch an interactive Hive cluster from the Amazon EMR console and the CLI.

Using Hive interactively is a great way to test query performance and tune your application. After you have established a set of Hive commands that will run on a regular basis, consider creating a Hive script that Amazon EMR can run for you.

> **Warning**
>
> Amazon EMR read or write operations on an DynamoDB table count against your established provisioned throughput, potentially increasing the frequency of provisioned throughput exceptions. For large requests, Amazon EMR implements retries with exponential backoff to manage the request load on the DynamoDB table. Running Amazon EMR jobs concurrently with other traffic may cause you to exceed the allocated provisioned throughput level. You can monitor this by checking the **ThrottleRequests** metric in Amazon CloudWatch. If the request load is too high, you can relaunch the cluster and set the Read percent setting or Write percent setting to a lower value to throttle the Amazon

EMR operations. For information about DynamoDB throughput settings, see Provisioned throughput.
If a table is configured for On-Demand mode, you should change the table back to provisioned mode before running an export or import operation. Pipelines need a throughput ratio in order to calculate resources to use from a DynamoDBtable. On-demand mode removes provisioned throughput. To provision throughput capacity, you can use Amazon CloudWatch Events metrics to evaluate the aggregate throughput that a table has used.

Topics

Set up a Hive table to run Hive commands

Apache Hive is a data warehouse application you can use to query data contained in Amazon EMR clusters using a SQL-like language. For more information about Hive, see http://hive.apache.org/.

The following procedure assumes you have already created a cluster and specified an Amazon EC2 key pair. To learn how to get started creating clusters, see Getting started with Amazon EMR in the *Amazon EMR Management Guide*.

Configure Hive to use MapReduce

When you use Hive on Amazon EMR to query DynamoDB tables, errors can occur if Hive uses the default execution engine, Tez. For this reason, when you create a cluster with Hive that integrates with DynamoDB as described in this section, we recommend that you use a configuration classification that sets Hive to use MapReduce. For more information, see Configure applications.

The following snippet shows the configuration classification and property to use to set MapReduce as the execution engine for Hive:

```
[
                {
                    "Classification": "hive-site",
                    "Properties": {
```

```
                    "hive.execution.engine": "mr"
                }
            }
        ]
```

To run Hive commands interactively

1. Connect to the master node. For more information, see [Connect to the master node using SSH](#) in the *Amazon EMR Management Guide*.
2. At the command prompt for the current master node, type `hive`.

   You should see a hive prompt: `hive>`
3. Enter a Hive command that maps a table in the Hive application to the data in DynamoDB. This table acts as a reference to the data stored in Amazon DynamoDB; the data is not stored locally in Hive and any queries using this table run against the live data in DynamoDB, consuming the table's read or write capacity every time a command is run. If you expect to run multiple Hive commands against the same dataset, consider exporting it first.

   The following shows the syntax for mapping a Hive table to a DynamoDB table.

   ```
   CREATE EXTERNAL TABLE hive_tablename
    (hive_column1_name column1_datatype, hive_column2_name column2_datatype...)
   STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
   TBLPROPERTIES ("dynamodb.table.name" = "dynamodb_tablename",
   "dynamodb.column.mapping" =
    "hive_column1_name:dynamodb_attribute1_name,hive_column2_name:dynamodb_attribute2_name..."
   ```

   When you create a table in Hive from DynamoDB, you must create it as an external table using the keyword `EXTERNAL`. The difference between external and internal tables is that the data in internal tables is deleted when an internal table is dropped. This is not the desired behavior when connected to Amazon DynamoDB, and thus only external tables are supported.

   For example, the following Hive command creates a table named *hivetable1* in Hive that references the DynamoDB table named *dynamodbtable1*. The DynamoDB table *dynamodbtable1* has a hash-and-range primary key schema. The hash key element is `name` (string type), the range key element is `year` (numeric type), and each item has an attribute value for `holidays` (string set type).

```
CREATE EXTERNAL TABLE hivetable1 (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");
```

Line 1 uses the HiveQL `CREATE EXTERNAL TABLE` statement. For *hivetable1*, you need to establish a column for each attribute name-value pair in the DynamoDB table, and provide the data type. These values are not case-sensitive, and you can give the columns any name (except reserved words).

Line 2 uses the `STORED BY` statement. The value of `STORED BY` is the name of the class that handles the connection between Hive and DynamoDB. It should be set to `'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'`.

Line 3 uses the `TBLPROPERTIES` statement to associate "hivetable1" with the correct table and schema in DynamoDB. Provide `TBLPROPERTIES` with values for the `dynamodb.table.name` parameter and `dynamodb.column.mapping` parameter. These values *are* case-sensitive.

Note

All DynamoDB attribute names for the table must have corresponding columns in the Hive table. Depending on your Amazon EMR version, the following scenarios occur if the one-to-one mapping does not exist:

- On Amazon EMR version 5.27.0 and later, the connector has validations that ensure a one-to-one mapping between DynamoDB attribute names and columns in the Hive table. An error will occur if the one-to-one mapping does not exist.
- On Amazon EMR version 5.26.0 and earlier, the Hive table won't contain the name-value pair from DynamoDB. If you do not map the DynamoDB primary key attributes, Hive generates an error. If you do not map a non-primary key attribute, no error is generated, but you won't see the data in the Hive table. If the data types do not match, the value is null.

Then you can start running Hive operations on *hivetable1*. Queries run against *hivetable1* are internally run against the DynamoDB table *dynamodbtable1* of your DynamoDB account, consuming read or write units with each execution.

When you run Hive queries against a DynamoDB table, you need to ensure that you have provisioned a sufficient amount of read capacity units.

For example, suppose that you have provisioned 100 units of read capacity for your DynamoDB table. This will let you perform 100 reads, or 409,600 bytes, per second. If that table contains 20GB of data (21,474,836,480 bytes), and your Hive query performs a full table scan, you can estimate how long the query will take to run:

21,474,836,480 / 409,600 = 52,429 seconds = 14.56 hours

The only way to decrease the time required would be to adjust the read capacity units on the source DynamoDB table. Adding more Amazon EMR nodes will not help.

In the Hive output, the completion percentage is updated when one or more mapper processes are finished. For a large DynamoDB table with a low provisioned read capacity setting, the completion percentage output might not be updated for a long time; in the case above, the job will appear to be 0% complete for several hours. For more detailed status on your job's progress, go to the Amazon EMR console; you will be able to view the individual mapper task status, and statistics for data reads. You can also log on to Hadoop interface on the master node and see the Hadoop statistics. This will show you the individual map task status and some data read statistics. For more information, see the following topics:

- Web interfaces hosted on the master node
- View the Hadoop web interfaces

For more information about sample HiveQL statements to perform tasks such as exporting or importing data from DynamoDB and joining tables, see Hive command examples for exporting, importing, and querying data in DynamoDB.

To cancel a Hive request

When you execute a Hive query, the initial response from the server includes the command to cancel the request. To cancel the request at any time in the process, use the **Kill Command** from the server response.

1. Enter `Ctrl+C` to exit the command line client.

2. At the shell prompt, enter the **Kill Command** from the initial server response to your request.

   Alternatively, you can run the following command from the command line of the master node to kill the Hadoop job, where *job-id* is the identifier of the Hadoop job and can be retrieved from the Hadoop user interface.

```
hadoop job -kill job-id
```

Data types for Hive and DynamoDB

The following table shows the available Hive data types, the default DynamoDB type that they correspond to, and the alternate DynamoDB types that they can also map to.

Hive type	Default DynamoDB type	Alternate DynamoDB type(s)
string	string (S)	
bigint or double	number (N)	
binary	binary (B)	
boolean	boolean (BOOL)	
array	list (L)	number set (NS), string set (SS), or binary set (BS)
map<string,string>	item	map (M)
map<string,?>	map (M)	
	null (NULL)	

If you want to write your Hive data as a corresponding alternate DynamoDB type, or if your DynamoDB data contains attribute values of an alternate DynamoDB type, you can specify the column and the DynamoDB type with the `dynamodb.type.mapping` parameter. The following example shows the syntax for specifying an alternate type mapping.

```
CREATE EXTERNAL TABLE hive_tablename (hive_column1_name column1_datatype,
 hive_column2_name column2_datatype...)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodb_tablename",
"dynamodb.column.mapping" =
 "hive_column1_name:dynamodb_attribute1_name,hive_column2_name:dynamodb_attribute2_name...",
"dynamodb.type.mapping" = "hive_column1_name:dynamodb_attribute1_datatype");
```

The type mapping parameter is optional, and only has to be specified for the columns that use alternate types.

For example, the following Hive command creates a table named `hivetable2` that references the DynamoDB table `dynamodbtable2`. It is similar to `hivetable1`, except that it maps the `col3` column to the string set (SS) type.

```
CREATE EXTERNAL TABLE hivetable2 (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbtable2",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays",
"dynamodb.type.mapping" = "col3:SS");
```

In Hive, `hivetable1` and `hivetable2` are identical. However, when data from those tables are written to their corresponding DynamoDB tables, `dynamodbtable1` will contain lists, while `dynamodbtable2` will contain string sets.

If you want to write Hive `null` values as attributes of DynamoDB `null` type, you can do so with the `dynamodb.null.serialization` parameter. The following example shows the syntax for specifying `null` serialization.

```
CREATE EXTERNAL TABLE hive_tablename (hive_column1_name column1_datatype,
 hive_column2_name column2_datatype...)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodb_tablename",
"dynamodb.column.mapping" =
 "hive_column1_name:dynamodb_attribute1_name,hive_column2_name:dynamodb_attribute2_name...",
```

```
"dynamodb.null.serialization" = "true");
```

The null serialization parameter is optional, and is set to `false` if not specified. Note that DynamoDB `null` attributes are read as `null` values in Hive regardless of the parameter setting. Hive collections with `null` values can be written to DynamoDB only if the null serialization parameter is specified as `true`. Otherwise, a Hive error occurs.

The bigint type in Hive is the same as the Java long type, and the Hive double type is the same as the Java double type in terms of precision. This means that if you have numeric data stored in DynamoDB that has precision higher than is available in the Hive datatypes, using Hive to export, import, or reference the DynamoDB data could lead to a loss in precision or a failure of the Hive query.

Exports of the binary type from DynamoDB to Amazon Simple Storage Service (Amazon S3) or HDFS are stored as a Base64-encoded string. If you are importing data from Amazon S3 or HDFS into the DynamoDB binary type, it should be encoded as a Base64 string.

Hive options

You can set the following Hive options to manage the transfer of data out of Amazon DynamoDB. These options only persist for the current Hive session. If you close the Hive command prompt and reopen it later on the cluster, these settings will have returned to the default values.

Hive options	Description
`dynamodb.throughput.read.percent`	Set the rate of read operations to keep your DynamoDB provisioned throughput rate in the allocated range for your table. The value is between `0.1` and `1.5`, inclusively. The value of 0.5 is the default read rate, which means that Hive will attempt to consume half of the read provisioned throughout resources in the table. Increasing this value above 0.5 increases the read request rate. Decreasing it below 0.5 decreases the read request rate. This read rate is approximate. The actual read rate will depend on factors such as whether there is a uniform distribution of keys in DynamoDB.

Hive options	Description
	If you find your provisioned throughput is frequently exceeded by the Hive operation, or if live read traffic is being throttled too much, then reduce this value below `0.5`. If you have enough capacity and want a faster Hive operation, set this value above `0.5`. You can also oversubscribe by setting it up to 1.5 if you believe there are unused input/output operations available.
`dynamodb.throughput.write.percent`	Set the rate of write operations to keep your DynamoDB provisioned throughput rate in the allocated range for your table. The value is between `0.1` and `1.5`, inclusively. The value of 0.5 is the default write rate, which means that Hive will attempt to consume half of the write provisioned throughout resources in the table. Increasing this value above 0.5 increases the write request rate. Decreasing it below 0.5 decreases the write request rate. This write rate is approximate. The actual write rate will depend on factors such as whether there is a uniform distribution of keys in DynamoDB If you find your provisioned throughput is frequently exceeded by the Hive operation, or if live write traffic is being throttled too much, then reduce this value below `0.5`. If you have enough capacity and want a faster Hive operation, set this value above `0.5`. You can also oversubscribe by setting it up to 1.5 if you believe there are unused input/output operations available or this is the initial data upload to the table and there is no live traffic yet.

Hive options	Description
dynamodb.endpoint	Specify the endpoint for the DynamoDB service. For more information about the available DynamoDB endpoints, see Regions and endpoints.
dynamodb.max.map.tasks	Specify the maximum number of map tasks when reading data from DynamoDB. This value must be equal to or greater than 1.
dynamodb.retry.duration	Specify the number of minutes to use as the timeout duration for retrying Hive commands. This value must be an integer equal to or greater than 0. The default timeout duration is two minutes.

These options are set using the SET command as shown in the following example.

```
SET dynamodb.throughput.read.percent=1.0;

INSERT OVERWRITE TABLE s3_export SELECT *
FROM hiveTableName;
```

Hive command examples for exporting, importing, and querying data in DynamoDB

The following examples use Hive commands to perform operations such as exporting data to Amazon S3 or HDFS, importing data to DynamoDB, joining tables, querying tables, and more.

Operations on a Hive table reference data stored in DynamoDB. Hive commands are subject to the DynamoDB table's provisioned throughput settings, and the data retrieved includes the data written to the DynamoDB table at the time the Hive operation request is processed by DynamoDB. If the data retrieval process takes a long time, some data returned by the Hive command may have been updated in DynamoDB since the Hive command began.

Hive commands `DROP TABLE` and `CREATE TABLE` only act on the local tables in Hive and do not create or drop tables in DynamoDB. If your Hive query references a table in DynamoDB, that table must already exist before you run the query. For more information about creating and deleting tables in DynamoDB, see Working with tables in DynamoDB in the *Amazon DynamoDB Developer Guide*.

> **Note**
>
> When you map a Hive table to a location in Amazon S3, do not map it to the root path of the bucket, s3://mybucket, as this may cause errors when Hive writes the data to Amazon S3. Instead map the table to a subpath of the bucket, s3://mybucket/mypath.

Exporting data from DynamoDB

You can use Hive to export data from DynamoDB.

To export a DynamoDB table to an Amazon S3 bucket

- Create a Hive table that references data stored in DynamoDB. Then you can call the INSERT OVERWRITE command to write the data to an external directory. In the following example, *s3://bucketname/path/subpath/* is a valid path in Amazon S3. Adjust the columns and datatypes in the CREATE command to match the values in your DynamoDB. You can use this to create an archive of your DynamoDB data in Amazon S3.

```
CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");


INSERT OVERWRITE DIRECTORY 's3://bucketname/path/subpath/' SELECT *
FROM hiveTableName;
```

To export a DynamoDB table to an Amazon S3 bucket using formatting

- Create an external table that references a location in Amazon S3. This is shown below as s3_export. During the CREATE call, specify row formatting for the table. Then, when you use INSERT OVERWRITE to export data from DynamoDB to s3_export, the data is written out in the specified format. In the following example, the data is written out as comma-separated values (CSV).

```
CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");


CREATE EXTERNAL TABLE s3_export(a_col string, b_col bigint, c_col array<string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 's3://bucketname/path/subpath/';

INSERT OVERWRITE TABLE s3_export SELECT *
FROM hiveTableName;
```

To export a DynamoDB table to an Amazon S3 bucket without specifying a column mapping

- Create a Hive table that references data stored in DynamoDB. This is similar to the preceding example, except that you are not specifying a column mapping. The table must have exactly one column of type `map<string, string>`. If you then create an `EXTERNAL` table in Amazon S3 you can call the `INSERT OVERWRITE` command to write the data from DynamoDB to Amazon S3. You can use this to create an archive of your DynamoDB data in Amazon S3. Because there is no column mapping, you cannot query tables that are exported this way. Exporting data without specifying a column mapping is available in Hive 0.8.1.5 or later, which is supported on Amazon EMR AMI 2.2.*x* and later.

```
CREATE EXTERNAL TABLE hiveTableName (item map<string,string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbtable1");

CREATE EXTERNAL TABLE s3TableName (item map<string, string>)
```

```
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n'
LOCATION 's3://bucketname/path/subpath/';

INSERT OVERWRITE TABLE s3TableName SELECT *
FROM hiveTableName;
```

To export a DynamoDB table to an Amazon S3 bucket using data compression

- Hive provides several compression codecs you can set during your Hive session. Doing so causes the exported data to be compressed in the specified format. The following example compresses the exported files using the Lempel-Ziv-Oberhumer (LZO) algorithm.

```
SET hive.exec.compress.output=true;
SET io.seqfile.compression.type=BLOCK;
SET mapred.output.compression.codec = com.hadoop.compression.lzo.LzopCodec;


CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");


CREATE EXTERNAL TABLE lzo_compression_table (line STRING)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n'
LOCATION 's3://bucketname/path/subpath/';

INSERT OVERWRITE TABLE lzo_compression_table SELECT *
FROM hiveTableName;
```

The available compression codecs are:

- org.apache.hadoop.io.compress.GzipCodec
- org.apache.hadoop.io.compress.DefaultCodec
- com.hadoop.compression.lzo.LzoCodec
- com.hadoop.compression.lzo.LzopCodec
- org.apache.hadoop.io.compress.BZip2Codec

- org.apache.hadoop.io.compress.SnappyCodec

To export a DynamoDB table to HDFS

- Use the following Hive command, where *hdfs:///directoryName* is a valid HDFS path and *hiveTableName* is a table in Hive that references DynamoDB. This export operation is faster than exporting a DynamoDB table to Amazon S3 because Hive 0.7.1.1 uses HDFS as an intermediate step when exporting data to Amazon S3. The following example also shows how to set `dynamodb.throughput.read.percent` to 1.0 in order to increase the read request rate.

```
CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

SET dynamodb.throughput.read.percent=1.0;

INSERT OVERWRITE DIRECTORY 'hdfs:///directoryName' SELECT * FROM hiveTableName;
```

  You can also export data to HDFS using formatting and compression as shown above for the export to Amazon S3. To do so, simply replace the Amazon S3 directory in the examples above with an HDFS directory.

To read non-printable UTF-8 character data in Hive

- You can read and write non-printable UTF-8 character data with Hive by using the `STORED AS SEQUENCEFILE` clause when you create the table. A SequenceFile is Hadoop binary file format; you need to use Hadoop to read this file. The following example shows how to export data from DynamoDB into Amazon S3. You can use this functionality to handle non-printable UTF-8 encoded characters.

```
CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbtable1",
```

```
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");


CREATE EXTERNAL TABLE s3_export(a_col string, b_col bigint, c_col array<string>)
STORED AS SEQUENCEFILE
LOCATION 's3://bucketname/path/subpath/';

INSERT OVERWRITE TABLE s3_export SELECT *
FROM hiveTableName;
```

Importing data to DynamoDB

When you write data to DynamoDB using Hive you should ensure that the number of write capacity units is greater than the number of mappers in the cluster. For example, clusters that run on m1.xlarge EC2 instances produce 8 mappers per instance. In the case of a cluster that has 10 instances, that would mean a total of 80 mappers. If your write capacity units are not greater than the number of mappers in the cluster, the Hive write operation may consume all of the write throughput, or attempt to consume more throughput than is provisioned. For more information about the number of mappers produced by each EC2 instance type, see Configure Hadoop.

The number of mappers in Hadoop are controlled by the input splits. If there are too few splits, your write command might not be able to consume all the write throughput available.

If an item with the same key exists in the target DynamoDB table, it is overwritten. If no item with the key exists in the target DynamoDB table, the item is inserted.

To import a table from Amazon S3 to DynamoDB

- You can use Amazon EMR (Amazon EMR) and Hive to write data from Amazon S3 to DynamoDB.

```
CREATE EXTERNAL TABLE s3_import(a_col string, b_col bigint, c_col array<string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 's3://bucketname/path/subpath/';

CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");
```

```
INSERT OVERWRITE TABLE hiveTableName SELECT * FROM s3_import;
```

To import a table from an Amazon S3 bucket to DynamoDB without specifying a column mapping

- Create an EXTERNAL table that references data stored in Amazon S3 that was previously exported from DynamoDB. Before importing, ensure that the table exists in DynamoDB and that it has the same key schema as the previously exported DynamoDB table. In addition, the table must have exactly one column of type `map<string, string>`. If you then create a Hive table that is linked to DynamoDB, you can call the `INSERT OVERWRITE` command to write the data from Amazon S3 to DynamoDB. Because there is no column mapping, you cannot query tables that are imported this way. Importing data without specifying a column mapping is available in Hive 0.8.1.5 or later, which is supported on Amazon EMR AMI 2.2.3 and later.

```
CREATE EXTERNAL TABLE s3TableName (item map<string, string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n'
LOCATION 's3://bucketname/path/subpath/';

CREATE EXTERNAL TABLE hiveTableName (item map<string,string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbtable1");

INSERT OVERWRITE TABLE hiveTableName SELECT *
FROM s3TableName;
```

To import a table from HDFS to DynamoDB

- You can use Amazon EMR and Hive to write data from HDFS to DynamoDB.

```
CREATE EXTERNAL TABLE hdfs_import(a_col string, b_col bigint, c_col array<string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 'hdfs:///directoryName';

CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
```

```
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

INSERT OVERWRITE TABLE hiveTableName SELECT * FROM hdfs_import;
```

Querying data in DynamoDB

The following examples show the various ways you can use Amazon EMR to query data stored in DynamoDB.

To find the largest value for a mapped column (max)

- Use Hive commands like the following. In the first command, the CREATE statement creates a Hive table that references data stored in DynamoDB. The SELECT statement then uses that table to query data stored in DynamoDB. The following example finds the largest order placed by a given customer.

```
CREATE EXTERNAL TABLE hive_purchases(customerId bigint, total_cost double,
 items_purchased array<String>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "Purchases",
"dynamodb.column.mapping" =
 "customerId:CustomerId,total_cost:Cost,items_purchased:Items");

SELECT max(total_cost) from hive_purchases where customerId = 717;
```

To aggregate data using the GROUP BY clause

- You can use the GROUP BY clause to collect data across multiple records. This is often used with an aggregate function such as sum, count, min, or max. The following example returns a list of the largest orders from customers who have placed more than three orders.

```
CREATE EXTERNAL TABLE hive_purchases(customerId bigint, total_cost double,
 items_purchased array<String>)
```

```
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "Purchases",
"dynamodb.column.mapping" =
 "customerId:CustomerId,total_cost:Cost,items_purchased:Items");

SELECT customerId, max(total_cost) from hive_purchases GROUP BY customerId HAVING
 count(*) > 3;
```

To join two DynamoDB tables

- The following example maps two Hive tables to data stored in DynamoDB. It then calls a join across those two tables. The join is computed on the cluster and returned. The join does not take place in DynamoDB. This example returns a list of customers and their purchases for customers that have placed more than two orders.

```
CREATE EXTERNAL TABLE hive_purchases(customerId bigint, total_cost double,
 items_purchased array<String>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "Purchases",
"dynamodb.column.mapping" =
 "customerId:CustomerId,total_cost:Cost,items_purchased:Items");

CREATE EXTERNAL TABLE hive_customers(customerId bigint, customerName string,
 customerAddress array<String>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "Customers",
"dynamodb.column.mapping" =
 "customerId:CustomerId,customerName:Name,customerAddress:Address");

Select c.customerId, c.customerName, count(*) as count from hive_customers c
JOIN hive_purchases p ON c.customerId=p.customerId
GROUP BY c.customerId, c.customerName HAVING count > 2;
```

To join two tables from different sources

- In the following example, Customer_S3 is a Hive table that loads a CSV file stored in Amazon S3 and hive_purchases is a table that references data in DynamoDB. The following example joins together customer data stored as a CSV file in Amazon S3 with order data stored in DynamoDB to return a set of data that represents orders placed by customers who have "Miller" in their name.

```
CREATE EXTERNAL TABLE hive_purchases(customerId bigint, total_cost double,
 items_purchased array<String>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "Purchases",
"dynamodb.column.mapping" =
 "customerId:CustomerId,total_cost:Cost,items_purchased:Items");

CREATE EXTERNAL TABLE Customer_S3(customerId bigint, customerName string,
 customerAddress array<String>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 's3://bucketname/path/subpath/';

Select c.customerId, c.customerName, c.customerAddress from
Customer_S3 c
JOIN hive_purchases p
ON c.customerid=p.customerid
where c.customerName like '%Miller%';
```

Note

In the preceding examples, the CREATE TABLE statements were included in each example for clarity and completeness. When running multiple queries or export operations against a given Hive table, you only need to create the table one time, at the beginning of the Hive session.

Optimizing performance for Amazon EMR operations in DynamoDB

Amazon EMR operations on a DynamoDB table count as read operations, and are subject to the table's provisioned throughput settings. Amazon EMR implements its own logic to try to balance the load on your DynamoDB table to minimize the possibility of exceeding your provisioned throughput. At the end of each Hive query, Amazon EMR returns information about the cluster used to process the query, including how many times your provisioned throughput was exceeded.

You can use this information, as well as CloudWatch metrics about your DynamoDB throughput, to better manage the load on your DynamoDB table in subsequent requests.

The following factors influence Hive query performance when working with DynamoDB tables.

Provisioned read capacity units

When you run Hive queries against a DynamoDB table, you need to ensure that you have provisioned a sufficient amount of read capacity units.

For example, suppose that you have provisioned 100 units of Read Capacity for your DynamoDB table. This will let you perform 100 reads, or 409,600 bytes, per second. If that table contains 20GB of data (21,474,836,480 bytes), and your Hive query performs a full table scan, you can estimate how long the query will take to run:

21,474,836,480 / 409,600 = 52,429 seconds = 14.56 hours

The only way to decrease the time required would be to adjust the read capacity units on the source DynamoDB table. Adding more nodes to the Amazon EMR cluster will not help.

In the Hive output, the completion percentage is updated when one or more mapper processes are finished. For a large DynamoDB table with a low provisioned Read Capacity setting, the completion percentage output might not be updated for a long time; in the case above, the job will appear to be 0% complete for several hours. For more detailed status on your job's progress, go to the Amazon EMR console; you will be able to view the individual mapper task status, and statistics for data reads.

You can also log on to Hadoop interface on the master node and see the Hadoop statistics. This shows you the individual map task status and some data read statistics. For more information, see Web interfaces hosted on the master node in the *Amazon EMR Management Guide*.

Read percent setting

By default, Amazon EMR manages the request load against your DynamoDB table according to your current provisioned throughput. However, when Amazon EMR returns information about your job that includes a high number of provisioned throughput exceeded responses, you can adjust the default read rate using the `dynamodb.throughput.read.percent` parameter when you set up the Hive table. For more information about setting the read percent parameter, see Hive options.

Write percent setting

By default, Amazon EMR manages the request load against your DynamoDB table according to your current provisioned throughput. However, when Amazon EMR returns information about your job that includes a high number of provisioned throughput exceeded responses, you can adjust the default write rate using the `dynamodb.throughput.write.percent` parameter when you set up the Hive table. For more information about setting the write percent parameter, see Hive options.

Retry duration setting

By default, Amazon EMR re-runs a Hive query if it has not returned a result within two minutes, the default retry interval. You can adjust this interval by setting the `dynamodb.retry.duration` parameter when you run a Hive query. For more information about setting the write percent parameter, see Hive options.

Number of map tasks

The mapper daemons that Hadoop launches to process your requests to export and query data stored in DynamoDB are capped at a maximum read rate of 1 MiB per second to limit the read capacity used. If you have additional provisioned throughput available on DynamoDB, you can improve the performance of Hive export and query operations by increasing the number of mapper daemons. To do this, you can either increase the number of EC2 instances in your cluster *or* increase the number of mapper daemons running on each EC2 instance.

You can increase the number of EC2 instances in a cluster by stopping the current cluster and re-launching it with a larger number of EC2 instances. You specify the number of EC2 instances in the **Configure EC2 Instances** dialog box if you're launching the cluster from the Amazon EMR console, or with the `--num-instances` option if you're launching the cluster from the CLI.

The number of map tasks run on an instance depends on the EC2 instance type. For more information about the supported EC2 instance types and the number of mappers each one provides, see Task configuration. There, you will find a "Task Configuration" section for each of the supported configurations.

Another way to increase the number of mapper daemons is to change the `mapreduce.tasktracker.map.tasks.maximum` configuration parameter of Hadoop to a higher value. This has the advantage of giving you more mappers without increasing either the number or the size of EC2 instances, which saves you money. A disadvantage is that setting this value too high can cause the EC2 instances in your cluster to run out of memory. To set `mapreduce.tasktracker.map.tasks.maximum`, launch the cluster and specify a value for `mapreduce.tasktracker.map.tasks.maximum` as a property of the mapred-site configuration classification. This is shown in the following example. For more information, see [Configure applications](#).

```
{
    "configurations": [
    {
        "classification": "mapred-site",
        "properties": {
            "mapred.tasktracker.map.tasks.maximum": "10"
        }
    }
    ]
}
```

Parallel data requests

Multiple data requests, either from more than one user or more than one application to a single table may drain read provisioned throughput and slow performance.

Process duration

Data consistency in DynamoDB depends on the order of read and write operations on each node. While a Hive query is in progress, another application might load new data into the DynamoDB table or modify or delete existing data. In this case, the results of the Hive query might not reflect changes made to the data while the query was running.

Avoid exceeding throughput

When running Hive queries against DynamoDB, take care not to exceed your provisioned throughput, because this will deplete capacity needed for your application's calls to `DynamoDB::Get`. To ensure that this is not occurring, you should regularly monitor the read volume and throttling on application calls to `DynamoDB::Get` by checking logs and monitoring metrics in Amazon CloudWatch.

Request time

Scheduling Hive queries that access a DynamoDB table when there is lower demand on the DynamoDB table improves performance. For example, if most of your application's users live in

San Francisco, you might choose to export daily data at 4 a.m. PST, when the majority of users are asleep, and not updating records in your DynamoDB database.

Time-based tables

If the data is organized as a series of time-based DynamoDB tables, such as one table per day, you can export the data when the table becomes no longer active. You can use this technique to back up data to Amazon S3 on an ongoing fashion.

Archived data

If you plan to run many Hive queries against the data stored in DynamoDB and your application can tolerate archived data, you may want to export the data to HDFS or Amazon S3 and run the Hive queries against a copy of the data instead of DynamoDB. This conserves your read operations and provisioned throughput.

Kinesis

Amazon EMR clusters can read and process Amazon Kinesis streams directly, using familiar tools in the Hadoop ecosystem such as Hive, Pig, MapReduce, the Hadoop Streaming API, and Cascading. You can also join real-time data from Amazon Kinesis with existing data on Amazon S3, Amazon DynamoDB, and HDFS in a running cluster. You can directly load the data from Amazon EMR to Amazon S3 or DynamoDB for post-processing activities. For information about Amazon Kinesis service highlights and pricing, see the Amazon Kinesis page.

What can I do with Amazon EMR and Amazon Kinesis integration?

Integration between Amazon EMR and Amazon Kinesis makes certain scenarios much easier; for example:

- **Streaming log analysis**–You can analyze streaming web logs to generate a list of top 10 error types every few minutes by region, browser, and access domain.
- **Customer engagement**–You can write queries that join clickstream data from Amazon Kinesis with advertising campaign information stored in a DynamoDB table to identify the most effective categories of ads that are displayed on particular websites.
- **Ad-hoc interactive queries**–You can periodically load data from Amazon Kinesis streams into HDFS and make it available as a local Impala table for fast, interactive, analytic queries.

Checkpointed analysis of Amazon Kinesis streams

Users can run periodic, batched analysis of Amazon Kinesis streams in what are called *iterations*. Because Amazon Kinesis stream data records are retrieved by using a sequence number, iteration boundaries are defined by starting and ending sequence numbers that Amazon EMR stores in a DynamoDB table. For example, when `iteration0` ends, it stores the ending sequence number in DynamoDB so that when the `iteration1` job begins, it can retrieve subsequent data from the stream. This mapping of iterations in stream data is called *checkpointing*. For more information, see Kinesis connector.

If an iteration was checkpointed and the job failed processing an iteration, Amazon EMR attempts to reprocess the records in that iteration.

Checkpointing is a feature that allows you to:

- Start data processing after a sequence number processed by a previous query that ran on same stream and logical name
- Re-process the same batch of data from Kinesis that was processed by an earlier query

To enable checkpointing, set the `kinesis.checkpoint.enabled` parameter to `true` in your scripts. Also, configure the following parameters:

Configuration setting	Description
kinesis.checkpoint.metastore.table.name	DynamoDB table name where checkpoint information will be stored
kinesis.checkpoint.metastore.hash.key.name	Hash key name for the DynamoDB table
kinesis.checkpoint.metastore.hash.range.name	Range key name for the DynamoDB table
kinesis.checkpoint.logical.name	A logical name for current processing
kinesis.checkpoint.iteration.no	Iteration number for processing associated with the logical name

Configuration setting	Description
kinesis.rerun.iteration.without.wait	Boolean value that indicates if a failed iteration can be rerun without waiting for timeout; the default is `false`

Provisioned IOPS recommendations for Amazon DynamoDB tables

The Amazon EMR connector for Amazon Kinesis uses the DynamoDB database as its backing for checkpointing metadata. You must create a table in DynamoDB before consuming data in an Amazon Kinesis stream with an Amazon EMR cluster in checkpointed intervals. The table must be in the same region as your Amazon EMR cluster. The following are general recommendations for the number of IOPS you should provision for your DynamoDB tables; let `j` be the maximum number of Hadoop jobs (with different logical name+iteration number combination) that can run concurrently and `s` be the maximum number of shards that any job will process:

For **Read Capacity Units**: `j*s/5`

For **Write Capacity Units**: `j*s`

Performance considerations

Amazon Kinesis shard throughput is directly proportional to the instance size of nodes in Amazon EMR clusters and record size in the stream. We recommend that you use m5.xlarge or larger instances on master and core nodes.

Schedule Amazon Kinesis analysis with Amazon EMR

When you are analyzing data on an active Amazon Kinesis stream, limited by timeouts and a maximum duration for any iteration, it is important that you run the analysis frequently to gather periodic details from the stream. There are multiple ways to execute such scripts and queries at periodic intervals; we recommend using AWS Data Pipeline for recurrent tasks like these. For more information, see AWS Data Pipeline PigActivity and AWS Data Pipeline HiveActivity in the *AWS Data Pipeline Developer Guide*.

Migrating Spark Kinesis connector to SDK 2.x for Amazon EMR 7.0

The AWS SDK provides a rich set of APIs and libraries to interact with AWS cloud computing services, such as managing credentials, connecting to S3 and Kinesis services. The Spark

Kinesis connector is used to consume data from Kinesis Data Streams, and the received data is transformed and processed in Spark's execution engine. Currently this connector is built on top of 1.x of AWS SDK and Kinesis-client-library (KCL).

As part of the AWS SDK 2.x migration, the Spark Kinesis connector is also updated accordingly to run with the SDK 2.x. In the Amazon EMR 7.0 release, Spark contains the SDK 2.x upgrade that is not yet available in the community version of Apache Spark. If you use the Spark Kinesis connector from a release that's lower than 7.0, you must migrate your application codes to run on SDK 2.x before you can migrate to Amazon EMR 7.0.

Migration guides

This section describes the steps to migrate an application to the upgraded Spark Kinesis connector. It includes guides to migrate to the Kinesis Client Library (KCL) 2.x, AWS credential providers, and AWS service clients in AWS SDK 2.x. For reference, it also includes a sample WordCount program that uses the Kinesis connector.

Topics

Migrating KCL from 1.x to 2.x

- **Metrics level and dimensions in `KinesisInputDStream`**

  When you instantiate a `KinesisInputDStream`, you can control the metrics level and dimensions for the stream. The following example demonstrates how you might customize these parameters with KCL 1.x:

```
import
 com.amazonaws.services.kinesis.clientlibrary.lib.worker.KinesisClientLibConfiguration
import com.amazonaws.services.kinesis.metrics.interfaces.MetricsLevel

val kinesisStream = KinesisInputDStream.builder
  .streamingContext(ssc)
```

```
  .streamName(streamName)
  .endpointUrl(endpointUrl)
  .regionName(regionName)
  .initialPosition(new Latest())
  .checkpointAppName(appName)
  .checkpointInterval(kinesisCheckpointInterval)
  .storageLevel(StorageLevel.MEMORY_AND_DISK_2)
  .metricsLevel(MetricsLevel.DETAILED)

 .metricsEnabledDimensions(KinesisClientLibConfiguration.DEFAULT_METRICS_ENABLED_DIMENSIONS.a
  .build()
```

In KCL 2.x, these config settings have different package names. To migrate to 2.x:

1. Change the import statements for `com.amazonaws.services.kinesis.clientlibrary.lib.worker.KinesisClientLibCon` and `com.amazonaws.services.kinesis.metrics.interfaces.MetricsLevel` to `software.amazon.kinesis.metrics.MetricsLevel` and `software.amazon.kinesis.metrics.MetricsUtil` respectively.

   ```
   // import com.amazonaws.services.kinesis.metrics.interfaces.MetricsLevel
   import software.amazon.kinesis.metrics.MetricsLevel

   // import
    com.amazonaws.services.kinesis.clientlibrary.lib.worker.KinesisClientLibConfiguration
   import software.amazon.kinesis.metrics.MetricsUtil
   ```

2. Replace the line `metricsEnabledDimensionsKinesisClientLibConfiguration.DEFAULT_METRICS_ENABL` with `metricsEnabledDimensionsSet(MetricsUtil.OPERATION_DIMENSION_NAME, MetricsUtil.SHARD_ID_DIMENSION_NAME)`

Following is an updated version of the `KinesisInputDStream` with customized metrics level and metrics dimensions:

```
import software.amazon.kinesis.metrics.MetricsLevel
import software.amazon.kinesis.metrics.MetricsUtil

val kinesisStream = KinesisInputDStream.builder
  .streamingContext(ssc)
```

```
  .streamName(streamName)
  .endpointUrl(endpointUrl)
  .regionName(regionName)
  .initialPosition(new Latest())
  .checkpointAppName(appName)
  .checkpointInterval(kinesisCheckpointInterval)
  .storageLevel(StorageLevel.MEMORY_AND_DISK_2)
  .metricsLevel(MetricsLevel.DETAILED)
  .metricsEnabledDimensions(Set(MetricsUtil.OPERATION_DIMENSION_NAME, 
 MetricsUtil.SHARD_ID_DIMENSION_NAME))
  .build()
```

- Message handler function in `KinesisInputDStream`

  When instantiating a `KinesisInputDStream`, you may also provide a “message handler function” that takes a Kinesis Record and returns a generic object T, in case you would like to use other data included in a Record such as partition key.

  In KCL 1.x, the message handler function signature is: `Record => T`, where Record is `com.amazonaws.services.kinesis.model.Record`. In KCL 2.x, the handler’s signature is changed to: `KinesisClientRecord => T`, where KinesisClientRecord is `software.amazon.kinesis.retrieval.KinesisClientRecord`.

  Following is an example of providing a message handler in KCL 1.x:

```
import com.amazonaws.services.kinesis.model.Record


def addFive(r: Record): Int = JavaUtils.bytesToString(r.getData).toInt + 5
val stream = KinesisInputDStream.builder
  .streamingContext(ssc)
  .streamName(streamName)
  .endpointUrl(endpointUrl)
  .regionName(regionName)
  .initialPosition(new Latest())
  .checkpointAppName(appName)
  .checkpointInterval(Seconds(10))
  .storageLevel(StorageLevel.MEMORY_ONLY)
  .buildWithMessageHandler(addFive)
```

  To migrate the message handler:

1. Change the import statement for `com.amazonaws.services.kinesis.model.Record` to `software.amazon.kinesis.retrieval.KinesisClientRecord`.

```
// import com.amazonaws.services.kinesis.model.Record
import software.amazon.kinesis.retrieval.KinesisClientRecord
```

2. Update the the method signature of the message handler.

```
//def addFive(r: Record): Int = JavaUtils.bytesToString(r.getData).toInt + 5
def addFive = (r: KinesisClientRecord) => JavaUtils.bytesToString(r.data()).toInt
 + 5
```

Following is an updated example of providing the message handler in KCL 2.x:

```
import software.amazon.kinesis.retrieval.KinesisClientRecord


def addFive = (r: KinesisClientRecord) => JavaUtils.bytesToString(r.data()).toInt + 5
val stream = KinesisInputDStream.builder
  .streamingContext(ssc)
  .streamName(streamName)
  .endpointUrl(endpointUrl)
  .regionName(regionName)
  .initialPosition(new Latest())
  .checkpointAppName(appName)
  .checkpointInterval(Seconds(10))
  .storageLevel(StorageLevel.MEMORY_ONLY)
  .buildWithMessageHandler(addFive)
```

For more information about migrating from KCL 1.x to 2.x, see Migrating Consumers from KCL 1.x to KCL 2.x.

Migrating AWS credentials providers from AWS SDK 1.x to 2.x

Credentials providers are used to obtain AWS credentials for interactions with AWS. There are several interface and class changes related to the credentials providers in SDK 2.x, which can be found here. Spark Kinesis connector has defined an interface (`org.apache.spark.streaming.kinesis.SparkAWSCredentials`) and implementation classes that returns 1.x version of AWS credential providers. These credentials providers are needed when initializing Kinesis clients. For instance, if you are using the method `SparkAWSCredentials.provider` in the applications, you would need to update codes to consume 2.x version of AWS credential providers.

Following is an example of using the credential providers in AWS SDK 1.x:

```
import org.apache.spark.streaming.kinesis.SparkAWSCredentials
import com.amazonaws.auth.AWSCredentialsProvider

val basicSparkCredentials = SparkAWSCredentials.builder
    .basicCredentials("accessKey", "secretKey")
    .build()

val credentialProvider = basicSparkCredentials.provider
assert(credentialProvider.isInstanceOf[AWSCredentialsProvider], "Type should be
 AWSCredentialsProvider")
```

To migrate to SDK 2.x:

1. Change the import statement for `com.amazonaws.auth.AWSCredentialsProvider` to `software.amazon.awssdk.auth.credentials.AwsCredentialsProvider`

```
//import com.amazonaws.auth.AWSCredentialsProvider
import software.amazon.awssdk.auth.credentials.AwsCredentialsProvider
```

2. Update the remaining codes that use this class.

```
import org.apache.spark.streaming.kinesis.SparkAWSCredentials
import software.amazon.awssdk.auth.credentials.AwsCredentialsProvider

val basicSparkCredentials = SparkAWSCredentials.builder
    .basicCredentials("accessKey", "secretKey")
    .build()

val credentialProvider = basicSparkCredentials.provider
assert (credentialProvider.isInstanceOf[AwsCredentialsProvider], "Type should be
 AwsCredentialsProvider")
```

Migrating AWS service clients from AWS SDK 1.x to 2.x

AWS service clients have different package names in 2.x (i.e. `software.amazon.awssdk`). whereas the SDK 1.x uses `com.amazonaws`. For more information about the client changes, see here. If you are using these service clients in the codes, you would need to migrate the clients accordingly.

Following is an example of creating a client in SDK 1.x:

```
import com.amazonaws.services.dynamodbv2.AmazonDynamoDBClient
import com.amazonaws.services.dynamodbv2.document.DynamoDB

AmazonDynamoDB ddbClient = AmazonDynamoDBClientBuilder.defaultClient();
AmazonDynamoDBClient ddbClient = new AmazonDynamoDBClient();
```

To migrate to 2.x:

1. Change the import statements for service clients. Take DynamoDB clients as an example. You would need to change `com.amazonaws.services.dynamodbv2.AmazonDynamoDBClient` or `com.amazonaws.services.dynamodbv2.document.DynamoDB` to `software.amazon.awssdk.services.dynamodb.DynamoDbClient`.

```
// import com.amazonaws.services.dynamodbv2.AmazonDynamoDBClient
// import com.amazonaws.services.dynamodbv2.document.DynamoDB
import software.amazon.awssdk.services.dynamodb.DynamoDbClient
```

2. Update the codes that initialize the clients

```
// AmazonDynamoDB ddbClient = AmazonDynamoDBClientBuilder.defaultClient();
// AmazonDynamoDBClient ddbClient = new AmazonDynamoDBClient();

DynamoDbClient ddbClient = DynamoDbClient.create();
DynamoDbClient ddbClient = DynamoDbClient.builder().build();
```

For more information about migrating AWS SDK from 1.x to 2.x, see What's different between the AWS SDK for Java 1.x and 2.x

Code examples for streaming applications

```
import java.net.URI
```

```
import software.amazon.awssdk.auth.credentials.DefaultCredentialsProvider
import software.amazon.awssdk.http.apache.ApacheHttpClient
import software.amazon.awssdk.services.kinesis.KinesisClient
import software.amazon.awssdk.services.kinesis.model.DescribeStreamRequest
import software.amazon.awssdk.regions.Region
import software.amazon.kinesis.metrics.{MetricsLevel, MetricsUtil}

import org.apache.spark.SparkConf
import org.apache.spark.storage.StorageLevel
import org.apache.spark.streaming.{Milliseconds, StreamingContext}
import org.apache.spark.streaming.dstream.DStream.toPairDStreamFunctions
import org.apache.spark.streaming.kinesis.KinesisInitialPositions.Latest
import org.apache.spark.streaming.kinesis.KinesisInputDStream


object KinesisWordCountASLSDKV2 {

  def main(args: Array[String]): Unit = {
    val appName = "demo-app"
    val streamName = "demo-kinesis-test"
    val endpointUrl = "https://kinesis.us-west-2.amazonaws.com"
    val regionName = "us-west-2"

    // Determine the number of shards from the stream using the low-level Kinesis
 Client
    // from the AWS Java SDK.
    val credentialsProvider = DefaultCredentialsProvider.create
    require(credentialsProvider.resolveCredentials() != null,
      "No AWS credentials found. Please specify credentials using one of the methods
 specified " +
        "in https://docs.aws.amazon.com/sdk-for-java/latest/developer-guide/
credentials.html")
    val kinesisClient = KinesisClient.builder()
      .credentialsProvider(credentialsProvider)
      .region(Region.US_WEST_2)
      .endpointOverride(URI.create(endpointUrl))
      .httpClientBuilder(ApacheHttpClient.builder())
      .build()
    val describeStreamRequest = DescribeStreamRequest.builder()
      .streamName(streamName)
      .build()
    val numShards = kinesisClient.describeStream(describeStreamRequest)
      .streamDescription
      .shards
```

```
      .size


    // In this example, we are going to create 1 Kinesis Receiver/input DStream for
 each shard.
    // This is not a necessity; if there are less receivers/DStreams than the number of
 shards,
    // then the shards will be automatically distributed among the receivers and each
 receiver
    // will receive data from multiple shards.
    val numStreams = numShards

    // Spark Streaming batch interval
    val batchInterval = Milliseconds(2000)

    // Kinesis checkpoint interval is the interval at which the DynamoDB is updated
 with information
    // on sequence number of records that have been received. Same as batchInterval for
 this
    // example.
    val kinesisCheckpointInterval = batchInterval

    // Setup the SparkConfig and StreamingContext
    val sparkConfig = new SparkConf().setAppName("KinesisWordCountASLSDKV2")
    val ssc = new StreamingContext(sparkConfig, batchInterval)

    // Create the Kinesis DStreams
    val kinesisStreams = (0 until numStreams).map { i =>
      KinesisInputDStream.builder
        .streamingContext(ssc)
        .streamName(streamName)
        .endpointUrl(endpointUrl)
        .regionName(regionName)
        .initialPosition(new Latest())
        .checkpointAppName(appName)
        .checkpointInterval(kinesisCheckpointInterval)
        .storageLevel(StorageLevel.MEMORY_AND_DISK_2)
        .metricsLevel(MetricsLevel.DETAILED)
        .metricsEnabledDimensions(Set(MetricsUtil.OPERATION_DIMENSION_NAME,
 MetricsUtil.SHARD_ID_DIMENSION_NAME))
        .build()
    }

    // Union all the streams
```

```
    val unionStreams = ssc.union(kinesisStreams)

    // Convert each line of Array[Byte] to String, and split into words
    val words = unionStreams.flatMap(byteArray => new String(byteArray).split(" "))

    // Map each word to a (word, 1) tuple so we can reduce by key to count the words
    val wordCounts = words.map(word => (word, 1)).reduceByKey(_ + _)

    // Print the first 10 wordCounts
    wordCounts.print()

    // Start the streaming context and await termination
    ssc.start()
    ssc.awaitTermination()
  }
}
```

Considerations when using the upgraded Spark Kinesis connector

- If your applications uses the `Kinesis-producer-library` with JDK version lower than 11, you may run into exceptions like `java.lang.NoClassDefFoundError: javax/xml/bind/DatatypeConverter`. This happens because EMR 7.0 comes with JDK 17 by default and J2EE modules have been removed from the standard libraries since Java 11+. This could be fixed by adding the following dependency in the pom file. Replace the library version with one as you see fit.

```
<dependency>
      <groupId>javax.xml.bind</groupId>
      <artifactId>jaxb-api</artifactId>
      <version>${jaxb-api.version}</version>
    </dependency>
```

- The Spark Kinesis connector jar can be found under this path after an EMR cluster is created: `/usr/lib/spark/connector/lib/`

S3DistCp (s3-dist-cp)

Apache DistCp is an open-source tool you can use to copy large amounts of data. *S3DistCp* is similar to DistCp, but optimized to work with AWS, particularly Amazon S3. The command for S3DistCp in Amazon EMR version 4.0 and later is `s3-dist-cp`, which you add as a step in a cluster or at the command line. Using S3DistCp, you can efficiently copy large amounts of data from Amazon S3 into HDFS where it can be processed by subsequent steps in your Amazon EMR cluster. You can also use S3DistCp to copy data between Amazon S3 buckets or from HDFS to Amazon S3. S3DistCp is more scalable and efficient for parallel copying large numbers of objects across buckets and across AWS accounts.

For specific commands that demonstrate the flexibility of S3DistCP in real-world scenarios, see Seven tips for using S3DistCp on the AWS Big Data blog.

Like DistCp, S3DistCp uses MapReduce to copy in a distributed manner. It shares the copy, error handling, recovery, and reporting tasks across several servers. For more information about the Apache DistCp open source project, see the DistCp guide in the Apache Hadoop documentation.

If S3DistCp is unable to copy some or all of the specified files, the cluster step fails and returns a non-zero error code. If this occurs, S3DistCp does not clean up partially copied files.

Important

S3DistCp does not support Amazon S3 bucket names that contain the underscore character.
S3DistCp does not support concatenation for Parquet files. Use PySpark instead. For more information, see Concatenating parquet files in Amazon EMR.
To avoid copy errors when using S3DistCP to copy a single file (instead of a directory) from S3 to HDFS, use Amazon EMR version 5.33.0 or later, or Amazon EMR version 6.3.0 or later.

S3DistCp options

Though similar to DistCp, S3DistCp supports a different set of options to change how it copies and compresses data.

When you call S3DistCp, you can specify the options described in the following table. The options are added to the step using the arguments list. Examples of the S3DistCp arguments are shown in the following table.

Option	Description	Required
`--src=LOCATION`	Location of the data to copy. This can be either an HDFS or Amazon S3 location.	Yes

Option	Description	Required
	Example: `--src=s3://amzn-s3-demo-bucket1/logs/j-3GYXXXXXX9IOJ/node` **Important** S3DistCp does not support Amazon S3 bucket names that contain the underscore character.	
`--dest=LOCATION`	Destination for the data. This can be either an HDFS or Amazon S3 location. Example: `--dest=hdfs:///output` **Important** S3DistCp does not support Amazon S3 bucket names that contain the underscore character.	Yes
`--srcPattern=PATTERN`	A regular expression that filters the copy operation to a subset of the data at `--src`. If neither `--srcPattern` nor `--groupBy` is specified, all data at `--src` is copied to `--dest`. If the regular expression argument contains special characters, such as an asterisk (*), either the regular expression or the entire `--args` string must be enclosed in single quotes ('). Example: `--srcPattern=.*daemons.*-hadoop-.*`	No

Option	Description	Required
--groupBy=PATTERN	A regular expression that causes S3DistCp to concatenate files that match the expression. For example, you could use this option to combine all of the log files written in one hour into a single file. The concatenated filename is the value matched by the regular expression for the grouping. Parentheses indicate how files should be grouped, with all of the items that match the parenthetical statement being combined into a single output file. If the regular expression does not include a parenthetical statement, the cluster fails on the S3DistCp step and return an error. If the regular expression argument contains special characters, such as an asterisk (*), either the regular expression or the entire `--args` string must be enclosed in single quotes ('). When `--groupBy` is specified, only files that match the specified pattern are copied. You do not need to specify `--groupBy` and `--srcPattern` at the same time. Example: `--groupBy=.*subnetid.*([0-9]+-[0-9]+-[0-9]+-[0-9]+).*`	No

Option	Description	Required
`--targetSize=SIZE`	The size, in mebibytes (MiB), of the files to create based on the `--groupBy` option. This value must be an integer. When `--targetSize` is set, S3DistCp attempts to match this size; the actual size of the copied files may be larger or smaller than this value. Jobs are aggregated based on the size of the data file, thus it is possible that the target file size will match the source data file size. If the files concatenated by `--groupBy` are larger than the value of `--targetSize`, they are broken up into part files, and named sequentially with a numeric value appended to the end. For example, a file concatenated into `myfile.gz` would be broken into parts as: `myfile0.gz`, `myfile1.gz`, etc. Example: `--targetSize=2`	No
`--appendToLastFile`	Specifies the behavior of S3DistCp when copying to files from Amazon S3 to HDFS which are already present. It appends new file data to existing files. If you use `--appendToLastFile` with `--groupBy`, new data is appended to files which match the same groups. This option also respects the `--targetSize` behavior when used with `--groupBy`.	No

Option	Description	Required
`--outputCodec=CODEC`	Specifies the compression codec to use for the copied files. This can take the values: `gzip`, `gz`, `lzo`, `snappy`, or `none`. You can use this option, for example, to convert input files compressed with Gzip into output files with LZO compression, or to uncompress the files as part of the copy operation. If you choose an output codec, the filename will be appended with the appropriate extension (e.g. for `gz` and `gzip`, the extension is `.gz`) If you do not specify a value for `--outputCodec` , the files are copied over with no change in their compression. Example: `--outputCodec=lzo`	No
`--s3ServerSideEncryption`	Ensures that the target data is transferred using SSL and automatically encrypted in Amazon S3 using an AWS service-side key. When retrieving data using S3DistCp, the objects are automatically unencrypted. If you attempt to copy an unencrypted object to an encryption-required Amazon S3 bucket, the operation fails. For more information, see Using data encryption. Example: `--s3ServerSideEncryption`	No
`--deleteOnSuccess`	If the copy operation is successful, this option causes S3DistCp to delete the copied files from the source location. This is useful if you are copying output files, such as log files, from one location to another as a scheduled task, and you don't want to copy the same files twice. Example: `--deleteOnSuccess`	No

Option	Description	Required
`--disableMultipartUpload`	Disables the use of multipart upload. Example: `--disableMultipartUpload`	No
`--multipartUploadChunkSize=SIZE`	The size, in MiB, of each part in an Amazon S3 multipart upload. S3DistCp uses multipart upload when it copies data larger than the `multipartUploadChunkSize` . To improve job performance, you can increase the size of each part. The default size is 128 MiB. Example: `--multipartUploadChunkSize=1000`	No
`--numberFiles`	Prepends output files with sequential numbers. The count starts at 0 unless a different value is specified by `--startingIndex` . Example: `--numberFiles`	No
`--startingIndex=INDEX`	Used with `--numberFiles` to specify the first number in the sequence. Example: `--startingIndex=1`	No
`--outputManifest=FILENAME`	Creates a text file, compressed with Gzip, that contains a list of all the files copied by S3DistCp. Example: `--outputManifest=manifest-1.gz`	No

Option	Description	Required
--previousManifest=PATH	Reads a manifest file that was created during a previous call to S3DistCp using the --outputManifest flag. When the --previousManifest flag is set, S3DistCp excludes the files listed in the manifest from the copy operation. If --outputManifest is specified along with --previousManifest , files listed in the previous manifest also appear in the new manifest file, although the files are not copied. Example: --previousManifest=/usr/bin/manifest-1.gz	No
--requirePreviousManifest	Requires a previous manifest created during a previous call to S3DistCp. If this is set to false, no error is generated when a previous manifest is not specified. The default is true.	No
--copyFromManifest	Reverses the behavior of --previousManifest to cause S3DistCp to use the specified manifest file as a list of files to copy, instead of a list of files to exclude from copying. Example: --copyFromManifest --previousManifest=/usr/bin/manifest-1.gz	No
--s3Endpoint=ENDPOINT	Specifies the Amazon S3 endpoint to use when uploading a file. This option sets the endpoint for both the source and destination. If not set, the default endpoint is s3.amazonaws.com . For a list of the Amazon S3 endpoints, see Regions and endpoints. Example: --s3Endpoint=s3.eu-west-1.amazonaws.com	No

Option	Description	Required
--storageClass=CLASS	The storage class to use when the destination is Amazon S3. Valid values are STANDARD and REDUCED_REDUNDANCY. If this option is not specified, S3DistCp tries to preserve the storage class. Example: `--storageClass=STANDARD`	No
--srcPrefixesFile=PATH	a text file in Amazon S3 (s3://), HDFS (hdfs:///) or local file system (file:/) that contains a list of `src` prefixes, one prefix per line. If `srcPrefixesFile` is provided, S3DistCp will not list the src path. Instead, it generates a source list as the combined result of listing all prefixes specified in this file. The relative path as compared to src path, instead of these prefixes, will be used to generate the destination paths. If `srcPattern` is also specified, it will be applied to the combined list results of the source prefixes to further filter the input. If `copyFromManifest` is used, objects in the manifest will be copied and `srcPrefixesFile` will be ignored. Example: `--srcPrefixesFile=PATH`	No

In addition to the options above, S3DistCp implements the Tool interface which means that it supports the generic options.

Adding S3DistCp as a step in a cluster

You can call S3DistCp by adding it as a step in your cluster. Steps can be added to a cluster at launch or to a running cluster using the console, CLI, or API. The following examples demonstrate adding an S3DistCp step to a running cluster. For more information on adding steps to a cluster, see Submit work to a cluster in the *Amazon EMR Management Guide*.

To add a S3DistCp step to a running cluster using the AWS CLI

For more information on using Amazon EMR commands in the AWS CLI, see the AWS CLI Command Reference.

- To add a step to a cluster that calls S3DistCp, pass the parameters that specify how S3DistCp should perform the copy operation as arguments.

  The following example copies daemon logs from Amazon S3 to `hdfs:///output`. In the following command:

  - `--cluster-id` specifies the cluster
  - `Jar` is the location of the S3DistCp JAR file. For an example of how to run a command on a cluster using command-runner.jar, see Submit a custom JAR step to run a script or command.
  - `Args` is a comma-separated list of the option name-value pairs to pass in to S3DistCp. For a complete list of the available options, see S3DistCp options.

  To add an S3DistCp copy step to a running cluster, put the following in a JSON file saved in Amazon S3 or your local file system as *`myStep.json`* for this example. Replace *`j-3GYXXXXXX9IOK`* with your cluster ID and replace *`mybucket`* with your Amazon S3 bucket name.

```
[
    {
        "Name":"S3DistCp step",
        "Args":["s3-dist-cp","--s3Endpoint=s3.amazonaws.com","--src=s3://mybucket/
logs/j-3GYXXXXXX9IOJ/node/","--dest=hdfs:///output","--srcPattern=.*[a-zA-Z,]+"],
        "ActionOnFailure":"CONTINUE",
        "Type":"CUSTOM_JAR",
        "Jar":"command-runner.jar"
    }
]
```

```
aws emr add-steps --cluster-id j-3GYXXXXXX9IOK --steps file://./myStep.json
```

Example Copy log files from Amazon S3 to HDFS

This example also illustrates how to copy log files stored in an Amazon S3 bucket into HDFS by adding a step to a running cluster. In this example the `--srcPattern` option is used to limit the data copied to the daemon logs.

To copy log files from Amazon S3 to HDFS using the `--srcPattern` option, put the following in a JSON file saved in Amazon S3 or your local file system as *`myStep.json`* for this example. Replace *`j-3GYXXXXXX9IOK`* with your cluster ID and replace *`mybucket`* with your Amazon S3 bucket name.

```
[
    {
        "Name":"S3DistCp step",
        "Args":["s3-dist-cp","--s3Endpoint=s3.amazonaws.com","--src=s3://mybucket/logs/
j-3GYXXXXXX9IOJ/node/","--dest=hdfs:///output","--srcPattern=.*daemons.*-hadoop-.*"],
        "ActionOnFailure":"CONTINUE",
        "Type":"CUSTOM_JAR",
        "Jar":"command-runner.jar"
    }
]
```

Cleaning up after failed S3DistCp jobs

If S3DistCp cannot copy some or all of the specified files, the command or cluster step fails and returns a non-zero error code. If this occurs, S3DistCp does not clean up partially copied files. You must delete them manually.

Partially copied files are saved to the HDFS `tmp` directory in sub-directories with the unique identifier of the S3DistCp job. You can find this ID in the standard output of the job.

For example, for an S3DistCp job with the ID `4b1c37bb-91af-4391-aaf8-46a6067085a6`, you can connect to the master node of the cluster and run the following command to view output files associated with the job.

```
hdfs dfs -ls /tmp/4b1c37bb-91af-4391-aaf8-46a6067085a6/output
```

The command returns a list of files similar to the following:

```
Found 8 items
```

```
-rw-r--r--   1 hadoop hadoop          0 2018-12-10 06:03 /tmp/4b1c37bb-91af-4391-
aaf8-46a6067085a6/output/_SUCCESS
-rw-r--r--   1 hadoop hadoop          0 2018-12-10 06:02 /tmp/4b1c37bb-91af-4391-
aaf8-46a6067085a6/output/part-r-00000
-rw-r--r--   1 hadoop hadoop          0 2018-12-10 06:02 /tmp/4b1c37bb-91af-4391-
aaf8-46a6067085a6/output/part-r-00001
-rw-r--r--   1 hadoop hadoop          0 2018-12-10 06:02 /tmp/4b1c37bb-91af-4391-
aaf8-46a6067085a6/output/part-r-00002
-rw-r--r--   1 hadoop hadoop          0 2018-12-10 06:03 /tmp/4b1c37bb-91af-4391-
aaf8-46a6067085a6/output/part-r-00003
-rw-r--r--   1 hadoop hadoop          0 2018-12-10 06:03 /tmp/4b1c37bb-91af-4391-
aaf8-46a6067085a6/output/part-r-00004
-rw-r--r--   1 hadoop hadoop          0 2018-12-10 06:03 /tmp/4b1c37bb-91af-4391-
aaf8-46a6067085a6/output/part-r-00005
-rw-r--r--   1 hadoop hadoop          0 2018-12-10 06:03 /tmp/4b1c37bb-91af-4391-
aaf8-46a6067085a6/output/part-r-00006
```

You can then run the following command to delete the directory and all contents.

```
hdfs dfs rm -rf /tmp/4b1c37bb-91af-4391-aaf8-46a6067085a6
```

Run commands and scripts on an Amazon EMR cluster

This topic covers how to run a command or a script as a step on your cluster. Running a command or script as a step is one of the many ways you can Submit work to a cluster and is useful in the following situations:

- When you don't have SSH access to your Amazon EMR cluster
- When you want to run a bash or shell command to troubleshoot your cluster

You can run a script either when you create a cluster or when your cluster is in the `WAITING` state. To run a script before step processing begins, you use a bootstrap action instead. For more information about bootstrap actions, see Create bootstrap actions to install additional software in the *Amazon EMR Management Guide*.

Amazon EMR provides the following tools to help you run scripts, commands, and other on-cluster programs. You can invoke both tools using the Amazon EMR management console or the AWS CLI.

`command-runner.jar`

Located on the Amazon EMR AMI for your cluster. You can use `command-runner.jar` to run commands on your cluster. You specify `command-runner.jar` without using its full path.

`script-runner.jar`

Hosted on Amazon S3 at `s3://`*`<region>`*`.elasticmapreduce/libs/script-runner/script-runner.jar` where *`<region>`* is the Region in which your Amazon EMR cluster resides. You can use `script-runner.jar` to run scripts saved locally or on Amazon S3 on your cluster. You must specify the full URI of `script-runner.jar` when you submit a step.

Submit a custom JAR step to run a script or command

The following AWS CLI examples illustrate some common use cases of `command-runner.jar` and `script-runner.jar` on Amazon EMR.

Example : Running a command on a cluster using `command-runner.jar`

When you use `command-runner.jar`, you specify commands, options, and values in your step's list of arguments.

The following AWS CLI example submits a step to a running cluster that invokes `command-runner.jar`. The specified command in the `Args` list downloads a script called *`my-script.sh`* from Amazon S3 into the hadoop user home directory. The command then modifies the script's permissions and runs *`my-script.sh`*.

When you use the AWS CLI, the items in your `Args` list should be comma separated with no whitespace between list elements. For example, `Args=[example-command,example-option,"example option value"]` instead of `Args=[example-command, example-option, "example option value"]`.

```
aws emr add-steps \
--cluster-id j-2AXXXXXXGAPLF \
--steps Type=CUSTOM_JAR,Name="Download a script from S3, change its permissions, and
 run it",ActionOnFailure=CONTINUE,Jar=command-runner.jar,Args=[bash,-c,"aws s3 cp s3://
EXAMPLE-DOC-BUCKET/my-script.sh /home/hadoop; chmod u+x /home/hadoop/my-script.sh; cd /
home/hadoop; ./my-script.sh"]
```

Example : Running a script on a cluster using `script-runner.jar`

When you use `script-runner.jar`, you specify the script that you want to run in your step's list of arguments.

The following AWS CLI example submits a step to a running cluster that invokes `script-runner.jar`. In this case, the script called *`my-script.sh`* is stored on Amazon S3. You can also specify local scripts that are stored on the master node of your cluster.

```
aws emr add-steps \
--cluster-id j-2AXXXXXXGAPLF \
--steps Type=CUSTOM_JAR,Name="Run a script from S3 with script-
runner.jar",ActionOnFailure=CONTINUE,Jar=s3://us-west-2.elasticmapreduce/libs/script-
runner/script-runner.jar,Args=[s3://EXAMPLE-DOC-BUCKET/my-script.sh]
```

Other ways to use `command-runner.jar`

You can also use `command-runner.jar` to submit work to a cluster with tools such as `spark-submit` or `hadoop-streaming`. When you launch an application using `command-runner.jar`, you specify `CUSTOM_JAR` as the step type instead of using a value like `SPARK`, `STREAMING`, or `PIG`. Tool availability varies depending on which applications you've installed on the cluster.

The following example command uses `command-runner.jar` to submit a step using `spark-submit`. The `Args` list specifies `spark-submit` as the command, followed by the Amazon S3 URI of the Spark application *my-app.py* with arguments and values.

```
aws emr add-steps \
--cluster-id j-2AXXXXXXGAPLF \
--steps Type=CUSTOM_JAR,Name="Run spark-submit using command-
runner.jar",ActionOnFailure=CONTINUE,Jar=command-runner.jar,Args=[spark-submit,S3://
DOC-EXAMPLE-BUCKET/my-app.py,ArgName1,ArgValue1,ArgName2,ArgValue2]
```

The following table identifies additional tools that you can run using `command-runner.jar`.

Tool name	Description
`hadoop-streaming`	Submits an Hadoop streaming program. In the console and some SDKs, this is a streaming step.
`hive-script`	Runs a Hive script. In the console and SDKs, this is a Hive step.
`pig-script`	Runs a Pig script. In the console and SDKs, this is a Pig step.
`spark-submit`	Runs a Spark application. In the console, this is a Spark step.
`hadoop-lzo`	Runs the Hadoop LZO indexer on a directory.
`s3-dist-cp`	Distributed copy large amounts of data from Amazon S3 into HDFS. For more information, see S3DistCp (s3-dist-cp).

AWS Glossary

For the latest AWS terminology, see the AWS glossary in the *AWS Glossary Reference*.